



Delaney, Christopher Daniel (2021) *Multi-omics approaches to understanding Candida biofilms*. PhD thesis.

<http://theses.gla.ac.uk/82245/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University of Glasgow

Multi-omics Approaches to Understanding *Candida* Biofilms

Christopher Daniel Delaney (BSc, MSc)

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Medicine
College of Medical, Veterinary and Life Sciences
University of Glasgow

December 2020

Abstract

Candida biofilms are a substantial clinical and human health burden which are still underappreciated. Benefits afforded by morphogenic switching from planktonic to biofilm communities include resistance to antimicrobials in the host's immune system, and resilience to mechanical disruption, all of which complicate the treatment and management of infections. Biofilm formation in *Candida* spp. is influenced by numerous factors, including response to the host, pH, bacteria, and many other environmental factors. This is further complicated by inherent heterogeneity within candidal populations in respect to biofilm forming capabilities and its response to external stimuli. Biofilms are heterogeneous by their nature, formed of populations of yeast and hyphal cells in a consortium of morphological states. Many variables exist to influence biofilm heterogeneity in *Candida* spp. and can influence the clinical outcomes and observations. The overarching aim of this thesis was to explore factors that influences *C. albicans* biofilm formation using omics-based approaches.

We performed analysis on the microbiome derived from three distinct oral niches in denture stomatitis patients. Alpha and beta diversity measures were extrapolated and compared to identify perturbations in the microbiome that were related to either the oral hygiene or the *Candida* burden of the individuals. Correlation analysis between phyla, oral hygiene and fungal load were performed to identify significant relationships.

Secondly, we selected clinical isolates, from our Scottish candidaemia study cohort, which were deemed high and low biofilm isolates as determined by biomass assays. We assessed the biofilm forming capabilities in media supplemented with and without serum. These assays consisted of crystal violet biomass assay and measurements of hyphal length. We also utilised SEM to visualise the phenotypes of the high and low isolates with and without serum. HBF and LBF isolates were then grown at 90min, 4h and 24h in the presence and absence of serum before being submitted to RNA-Sequencing by Illumina. Differential expression analysis was performed using DESeq2 before over representation and gene set enrichment analysis. Cell wall proteomics on

high and low isolates was also performed to identify computational changes in the cell wall.

Thirdly, we performed metabolomic analysis both targeted and untargeted on the supernatants of HBF and LBF grown in the presence or absence of serum for 4 and 24h. Features were identified from the LC-MS peaks by PiMP and analysis of differentially abundant analytes between our isolates and growth conditions were performed. Functional analysis of the annotated analytes was then performed by pathway activity profiling.

Finally, utilising our identified differentially expressed features from metabolomic and transcriptomic analysis we submitted both datasets to integrative analysis. Using a combination of conceptual, joint pathway analysis by MetaboAnalyst and multivariate analysis using the MixOmics data integration package.

From our microbiome analysis we observed that the oral hygiene measures had no significant effect on the diversity or composition of the oral microbiome. *Candida* similarly had no impact on the alpha diversity of the oral microbiome. However, we did observe some relationship in the beta diversity which correlated with *Candida* load. Further investigation identified correlations of genus including *Lactobacillus* with the *Candida* load. From our high and low *Candida* candidaemia isolates we observed phenotypic switching of LBF in the presence of serum. We also found functional differences related to this phenotypic switching. The low biofilm response to serum included enrichment in fatty acid and aycI-coA metabolic pathways. Metabolomic analysis revealed changes in arachidonic acid metabolism in serum grown isolates and changes in the amino acid metabolism between LBF and HBF isolates. Integrating these data, we were able to observe overlaps in the metabolic reprogramming of *C. albicans* isolates in serum with joint pathway analysis confirming changes in the fatty acid metabolic response in both transcriptomic and metabolomic data. Multivariate analysis by sPLS-DA identified several highly covariate discriminatory features with and without the presence of serum. These included many genes of currently unknown function and a downregulation of specific genes in serum including zinc transport.

Through the application of transcriptomics and metabolomics we have demonstrated that these holistic methodologies are invaluable to biofilm research. We identified molecular processes and metabolomic reprogramming of *C. albicans* in response to the biofilm inducing stimulus of serum. We also highlight the current and potential benefits that integration of multiple omics data sets provides. Integration is not without its challenges, however, and we identify some key methodologies that could improve interpretability of omics datasets derived from microbial communities. As *Candida* spp. do not exist within a vacuum, and infectious disease aetiology is dependent on the interactions between fungal and bacterial species, understanding the mechanisms that govern these biofilm models will help us to identify important factors and potential therapeutic strategies.

Table of Contents

ABSTRACT	I
TABLES	VI
FIGURES	VI
ACKNOWLEDGEMENTS.....	IX
RELEVANT PUBLICATIONS	XI
ACCOMPANYING MATERIAL	XII
ABBREVIATIONS	XIII
1 CANDIDA BIOFILMS, THEIR CLINICAL SIGNIFICANCE AND ROLE IN INTERKINGDOM INTERACTIONS.....	1
1.1 INTRODUCTION	2
1.2 CANDIDA BIOFILMS	3
1.2.1 Adhesion and virulence	5
1.2.2 Biofilm matrix.....	6
1.2.3 Molecular basis and transcriptional regulation	7
1.2.4 Metabolism	8
1.3 BIOFILM HETEROGENEITY.....	9
1.3.1 Is heterogeneity clinically important?.....	9
1.3.2 How does heterogeneity impact antifungal treatment?.....	11
1.3.3 What drives biofilm heterogeneity?.....	12
1.4 POLYMICROBIALITY IN THE ORAL CAVITY	15
1.4.1 Candida and the oral cavity	16
1.4.2 Periodontal disease.....	17
1.4.3 Denture stomatitis	20
1.5 METHODOLOGICAL APPROACHES TO UNDERSTANDING CANDIDA AND MIXED-SPECIES BIOFILMS.....	22
1.5.1 The conventional and established approaches	22
1.5.2 The new and innovative approaches.....	25
1.5.3 Microbiome	25
1.5.4 The mycobiome	29
1.5.5 Transcriptome	31
1.5.6 Metabolome.....	33
1.5.7 Proteome.....	36
1.5.8 Integrative Analysis.....	38
1.6 CONCLUDING REMARKS AND AIMS.....	39
2 THE IMPACT OF CANDIDA AND ITS INTERACTIONS WITH THE MICROBIOTA AT THE DENTURE SURFACE.....	42
2.1 BACKGROUND	43
2.2 HYPOTHESIS AND AIMS	44
2.3 METHODS	45
2.3.1 In vitro denture cleansing study.....	45
2.3.2 In vivo denture hygiene study	48
2.3.3 Study participants	49
2.3.4 Clinical sample collection	49
2.3.5 DNA Isolation	50
2.3.6 Illumina Sequencing	51
2.3.7 Sequencing data analysis.....	51
2.3.8 Study design and statistical analysis.....	52
2.4 RESULTS.....	53
2.4.1 Influence of denture hygiene on in vitro biofilms.....	53
2.4.2 Influence of denture hygiene on the denture associated Candida.....	56
2.4.3 Influence of denture hygiene on the oral microbiome	58

2.4.4	<i>Influence of Candida load on hygiene and microbial communities</i>	64
2.5	DISCUSSION	69
2.6	HIGHLIGHTS.....	73
3	TRANSCRIPTOMIC PROFILING OF PHENOTYPICALLY DISTINCT <i>CANDIDA ALBICANS</i> CLINICAL ISOLATES	74
3.1	BACKGROUND AND INTRODUCTION	75
3.2	AIMS	76
3.3	METHODS	77
3.3.1	<i>Culture conditions and standardisation</i>	77
3.3.2	<i>Biofilm assay</i>	78
3.3.3	<i>RNA extraction and sequencing</i>	78
3.3.4	<i>RNA-Seq analysis</i>	80
3.3.5	<i>Electron microscopy</i>	81
3.3.6	<i>Cell wall proteomic analysis</i>	82
3.4	RESULTS.....	83
3.4.1	<i>Assessing the Biofilm Phenotype</i>	83
3.4.2	<i>RNA-Seq read quality</i>	85
3.4.3	<i>RNA-Seq multivariate analysis</i>	87
3.4.4	<i>Candida albicans differential expression in response to nutrient stress</i>	89
3.4.5	<i>Gene expression of strain and time dependant genes in the nutrient stress response of Candida albicans</i>	93
3.4.6	<i>Gene set enrichment analysis of strains in the presence of serum</i>	109
3.4.7	<i>Biomarker analysis</i>	111
3.4.8	<i>Cell Wall Proteomics</i>	114
3.5	DISCUSSION	118
3.6	HIGHLIGHTS.....	123
4	METABOLOMIC PROFILING OF PHENOTYPICALLY DISTINCT <i>CANDIDA ALBICANS</i> CLINICAL ISOLATES	124
4.1	BACKGROUND	125
4.2	AIMS	127
4.3	METHODS	128
4.3.1	<i>Biofilm Preparation</i>	128
4.3.2	<i>Data Acquisition and Processing</i>	129
4.3.3	<i>Statistical Analysis</i>	129
4.4	RESULTS.....	131
4.4.1	<i>Comparisons between media controls</i>	131
4.4.2	<i>Metabolomic profiling of Low and High Biofilm Forming Isolates</i>	133
4.4.3	<i>Multivariate analysis of C. albicans HBF and LBF</i>	138
4.4.4	<i>Profiling of targeted metabolites in isolates grown in serum vs RPMI</i>	140
4.4.5	<i>Identified amino acid metabolism</i>	142
4.4.6	<i>Identified Nucleotide metabolism</i>	146
4.4.7	<i>Carbohydrates</i>	146
4.4.8	<i>Identified lipid and fatty acids</i>	147
4.4.9	<i>Pathway activity profiling of C. albicans metabolic footprint</i>	149
4.5	DISCUSSION	156
4.6	HIGHLIGHTS.....	162
5	OMICS INTEGRATION AND DISCUSSION	163
5.1	INTRODUCTION	164
5.2	UNDERSTANDING THE BIOGEOGRAPHY OF BIOFILMS.....	164
5.3	OMICS APPROACHES TO <i>CANDIDA ALBICANS</i> BIOFILMS.....	167
5.4	INTEGRATION	168
5.4.1	<i>Conceptual integration</i>	170
5.4.2	<i>Pathway models</i>	174
5.4.3	<i>Statistical and multivariate methods</i>	178

5.4.4	<i>Data Integration Conclusions</i>	183
5.5	FINAL DISCUSSION AND CONCLUSIONS	183
REFERENCES		189

TABLES

TABLE 2.1 MAINTENANCE AND CULTURE CONDITIONS FOR ORAL DENTURE SPECIES	46
TABLE 2.2 PRIMERS FOR QPCR AND ILLUMINA SEQUENCING	50
TABLE 2.3 PATIENT DENTURE HYGIENE DEMOGRAPHICS	55
TABLE 3.1 CANDIDEMIA ISOLATES USED IN EACH INDIVIDUAL EXPERIMENTAL DESIGN	79
TABLE 3.2 RNA-SEQUENCING, QUALITY CONTROL AND ALIGNMENT STATISTICS	86
TABLE 3.3 DESEQ2 RESULTS OF THE TOP 10 UP AND TOP 10 DOWN REGULATED GENES IN HBF.....	97
TABLE 3.4 DESEQ2 RESULTS OF THE TOP 10 UP AND TOP 10 DOWN REGULATED GENES IN LBF.....	98
TABLE 3.5 TABLE OF IDENTIFIED CELL WALL PROTEINS AND THEIR PRESENCE/ABSENCE WITHIN EACH GROUP OF STRAINS	117
TABLE 4.1 FOLD CHANGES OF FCS METABOLITES COMPARED TO RPMI	132

FIGURES

FIGURE 1.1 SIMPLIFIED DEVELOPMENTAL CYCLE OF <i>C. ALBICANS</i> BIOFILM.....	5
FIGURE 1.2 MOLECULAR BASIS OF <i>C. ALBICANS</i> BIOFILM FORMATION	8
FIGURE 1.3 MUCOSAL AND MEDICAL DEVICES ASSOCIATED <i>CANDIDA</i> BIOFILM INFECTIONS	11
FIGURE 1.4 SCHEMATIC REPRESENTATION OF FACTORS DRIVING BIOFILM-ASSOCIATED ANTIFUNGAL TOLERANCE	12
FIGURE 1.5 MAXIMUM SCORING METABOLIC SUBNETWORK IN THE LBF-HBF NETWORK	14
FIGURE 1.6 FACTORS INFLUENCING BIOFILM FORMATION	15
FIGURE 2.1 TREATMENT REGIMEN FOR INVITRO ASSESSMENT OF DENTURE CLEANSING ON BACTERIAL AND FUNGAL LOADS.....	47
FIGURE 2.2 SAMPLE COLLECTION AND ANALYSIS WORKFLOW	48
FIGURE 2.3 QUANTITATIVE LIVE ASSESSMENT OF BACTERIAL AND <i>CANDIDA</i> LOAD FOLLOWING <i>IN VITRO</i> DENTURE HYGIENE	54
FIGURE 2.4 <i>CANDIDA</i> LOAD BETWEEN HYGIENE VARIABLES AND ORAL SITES	57
FIGURE 2.5 PCOA ORDINATION PLOT BASED ON PHYLOSEQ'S WEIGHTED UNIFRAC DISTANCES.....	59
FIGURE 2.6 MICROBIOME DIVERSITY MEASURES BETWEEN POOR, GOOD AND EXCELLENT ORAL HYGIENE	60
FIGURE 2.7 MICROBIOME DIVERSITY MEASURES BETWEEN DENTURE CLEANING ONCE OR TWICE.	61
FIGURE 2.8 MICROBIOME DIVERSITY MEASURES BETWEEN SLEEPING WITH OR NOT SLEEPING WITH DENTURES <i>IN SITU</i>	62
FIGURE 2.9 MEASURE OF TAXA IN SIGNIFICANTLY HIGHER ABUNDANCE BETWEEN PATIENT VARIABLES	63
FIGURE 2.10 NON-METRIC DIMENSIONAL SCALING OF OTU DATA BASED UPON THE BRAY-CURTIS DISTANCE MEASURE.....	65
FIGURE 2.11 MICROBIOME DIVERSITY MEASURES BETWEEN <i>CANDIDA</i> LOAD LOW, MEDIUM AND HIGH	66
FIGURE 2.12 CORRELATIONS OF VARIABLES WITH THE LEVELS OF SPECIFIC GENUS ON THE MUCOSAL, DENTURE AND DENTAL SURFACE	68
FIGURE 3.1 OVERVIEW OF SAMPLE PREPARATION, RNA-SEQUENCING, AND ANALYSIS APPROACHES	80
FIGURE 3.2 VARIATION IN <i>C. ALBICANS</i> BIOFILM FORMATION IN THE PRESENCE AND ABSENCE OF FCS.....	84
FIGURE 3.3 BIOMASS OF LBF ISOLATES GROWN IN RPMI AND RPMI SUPPLEMENTED WITH 25% SERUM	85
FIGURE 3.4 PRINCIPAL COMPONENTS ANALYSIS (PCA) OF <i>C. ALBICANS</i> TRANSCRIPTOME	88
FIGURE 3.5 SCREE PLOT OF PCA COMPONENTS	89
FIGURE 3.6 VOLCANO PLOT DEPICTING GROSS CHANGES OF <i>C. ALBICANS</i> GROWN IN RPMI VS SERUM	90
FIGURE 3.7 BAR PLOT DEPICTING THE DIFFERENTIALLY REGULATED GENE ONTOLOGY AND KEGG PATHWAYS.....	92
FIGURE 3.8 NUMBER OF UP OR DOWNREGULATED AT EACH TIME POINT FOR EACH STRAIN IN THE PRESENCE OF SERUM	93

FIGURE 3.9 VOLCANO PLOTS DEPICTING THE LEVELS OF GENE EXPRESSION AT EACH TIME AND WITHIN EITHER THE HBF OR LBF <i>C. ALBICANS</i> STRAIN	94
FIGURE 3.10 UPSET PLOTS DEPICTING THE OVERLAP OF DIFFERENTIALLY EXPRESSED GENES	95
FIGURE 3.11 CLUEGO ANALYSIS OF THE UPREGULATED GENES IN LBF IN PRESENCE OF SERUM AND THEIR ASSOCIATED PATHWAYS	101
FIGURE 3.12 CLUEGO ANALYSIS OF THE DOWNREGULATED GENES IN LBF IN PRESENCE OF SERUM AND THEIR ASSOCIATED PATHWAYS	103
FIGURE 3.13 CLUEGO ANALYSIS OF THE UPREGULATED GENES IN LBF AND HBF STRAINS IN THE PRESENCE OF SERUM AT 90MIN.	106
FIGURE 3.14 CLUEGO ANALYSIS OF THE UPREGULATED GENES IN LBF AND HBF STRAINS IN THE PRESENCE OF SERUM AT 4H	107
FIGURE 3.15 CLUEGO ANALYSIS OF THE UPREGULATED GENES IN LBF AND HBF STRAINS IN THE PRESENCE OF SERUM AT 4H	108
FIGURE 3.16 GENE SET ENRICHMENT ANALYSIS DEMONSTRATING DIFFERENTIALLY REGULATED PATHWAYS IN THE PRESENCE/ABSENCE OF SERUM IN HIGH AND LOW BIOFILM FORMERS AT 24H	110
FIGURE 3.17 SUPERVISED ANALYSIS OF HBF AND LBF STRAINS IN THE PRESENCE/ABSENCE OF SERUM USING PLS-DA.	113
FIGURE 3.18 CELL WALL PROTEOMICS OF HBF AND LBF <i>C. ALBICANS</i> ISOLATES.	116
FIGURE 4.1 SCHEMATIC DEPICTING OUR WORKFLOW	130
FIGURE 4.2 VOLCANO PLOTS DEPICTING DIFFERENTIALLY REPRESENTED METABOLITES IN OUR MEDIA CONTROLS.....	131
FIGURE 4.3 VOLCANO PLOTS SHOWING DIFFERENCES BETWEEN LBF AND HBF AT EACH RESPECTIVE TIMEPOINT IN BOTH RPMI AND RPMI+FCS	134
FIGURE 4.4 BAR CHART OF SIGNIFICANTLY CHANGING IDENTIFIED COMPOUNDS BETWEEN HBF AND LBF	135
FIGURE 4.5 PARTIAL LEAST SQUARES OF LBF AND HBF <i>C. ALBICANS</i> ISOLATES AT 24H.	137
FIGURE 4.6 PRINCIPAL COMPONENT ANALYSIS (PCA) OF METABOLOMICS SECRETOME/METABOLIC FOOTPRINT ANALYSIS OF <i>C. ALBICANS</i> HIGH AND LOW BIOFILMS IN THE PRESENCE OR ABSENCE OF SERUM AT 4H OR 24H.	139
FIGURE 4.7 GRID OR HEAT PLOT OF THE RELATIVE ABUNDANCE COMPARED TO THE MEDIA CONTROL OF THE TARGETED OR IDENTIFIED METABOLITES	141
FIGURE 4.8 IDENTIFIED METABOLITES BETWEEN RPMI AND SERUM IN <i>C. ALBICANS</i> ISOLATES.....	143
FIGURE 4.9 IDENTIFIED METABOLITES BETWEEN RPMI AND SERUM IN <i>C. ALBICANS</i> ISOLATES.....	144
FIGURE 4.10 IDENTIFIED METABOLITES BETWEEN RPMI AND SERUM IN <i>C. ALBICANS</i> ISOLATES....	148
FIGURE 4.11 . GRID PLOT OF THE LEVEL OF PATHWAY ACTIVATION ACCORDING TO PATHWAY ACTIVITY PROFILING (PAPI)	150
FIGURE 4.12 TOP 8 ACTIVATED PATHWAYS ACCORDING TO PAPI	151
FIGURE 4.13 SECOND TOP 4 ACTIVATED PATHWAYS ACCORDING TO PAPI.....	152
FIGURE 4.14 BAR CHART COMPARING THE RELATIVE ABUNDANCE OF ARACHIDONIC ACID.....	155
FIGURE 5.1 COMPARATIVE ANALYSIS OF THE ENRICHED KEGG TERMS SERUM IN BOTH LOW AND HIGH BIOFILM FORMING <i>C. ALBICANS</i> ISOLATES.	169
FIGURE 5.2 DIFFERENTIAL EXPRESSION OF GENES ASSOCIATED WITH KEY METABOLIC PATHWAYS.	173
FIGURE 5.3 INTEGRATED PATHWAY ENRICHMENT FROM <i>C. ALBICANS</i> TRANSCRIPTS AND METABOLOMES	176
FIGURE 5.4 SPARSE PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS (SPLS-DA) OF <i>C. ALBICANS</i> ISOLATES IN THE PRESENCE AND ABSENCE OF SERUM	181
FIGURE 5.5 IMPORTANT FEATURES AND CORRELATIONS BETWEEN METABOLITES AND TRANSCRIPTS	182

Acknowledgements

It seems only right to begin by thanking my supervisors Professor Gordon Ramage, Dr Chris Nile, Dr Marcello Riggio and Dr Karl Burgess. Gordon, you have provided numerous opportunities, support, and guidance throughout my PhD. It was because of you that I was able to undertake my PhD and because of your support that I was able to make it through.

I would also like to express my appreciation to my industrial sponsors GlaxoSmithKline, for their financial support that they have provided throughout this PhD. I also would like to thank David Bradshaw who always offered positivity, encouragement, and support.

I wish to thank all other members of the Oral Sciences and GDH, past and present, who have helped me along the way. Including Dr Leighann Sherry, Dr Ranjith Rajendran, Dr Ryan Kean, Dr Jason Brown, Dr Will McLean, Professor Shauna Culshaw, Dr David Lappin and Mr Steven Milligan. You have all supported, advised and shaped the great experience I had whilst undertaking my PhD at the GDH.

I would also like to thank all those that supported me from outside of the GDH, who I was able to reach out to for assistance or collaboration for the work that was undertaken. I would like to thank Dr Umer Ijaz for his help as well as Dr Konstantinos Gerasimidis for his expert advice. I would also like to thank Dr Carol Munro (University of Aberdeen, UK) for providing the cell wall proteomics data.

A big thank you to all the members of the GDH lab: Bryn Short, Mark Butcher, Tracy Young, Khawlah Albashaireh, Om Alkhair Alshanta, Abeer Alghamdi, William Johnston, Emily McKloud, Saeed Alqahtani and Sumaya Abrusrewil. Over the past four years you have all helped to make the GDH a great place to work.

I would like to thank me mam, me dad and me sister for supporting me to go and undertake a PhD and their continued encouragement throughout. You have been

fantastic, and I would not have been able to do it without you. Cheers times a million to Seb and Brynley for always being on the other end of the phone for a good ol' rant. Steph, the only person more relieved that these 4 years are over than me! Thank you for everything, for bearing with me, giving me support and cheering me on when I needed it the most.

List of publications

Seneviratne, C.J.; Suriyanarayanan, T.; Widyarman, A.S.; Lee, L.S.; Lau, M.; Ching, J.; **Delaney, C.**; Ramage, G. Multi-omics tools for studying microbial biofilms: current perspectives and future directions. *Crit Rev Microbiol* **2020**, *46*, 759-778, doi:10.1080/1040841X.2020.1828817.

Brown JL, **Delaney C**, Short B, Butcher MC, McKloud E, Williams C, Kean R, Ramage G. *Candida auris* Phenotypic Heterogeneity Determines Pathogenicity *In Vitro*. *mSphere*. 2020 Jun 24;5(3):e00371-20. doi: 10.1128/mSphere.00371-20. PMID: 32581078; PMCID: PMC7316489.

Short B, Brown J, **Delaney C**, Sherry L, Williams C, Ramage G, Kean R. *Candida auris* exhibits resilient biofilm characteristics in vitro: implications for environmental persistence. *J Hosp Infect*. 2019 Sep;103(1):92-96. doi: 10.1016/j.jhin.2019.06.006. Epub 2019 Jun 18. PMID: 31226270.

Delaney, C.; O'Donnell, L.E.; Kean, R.; Sherry, L.; Brown, J.L.; Calvert, G.; Nile, C.J.; Cross, L.; Bradshaw, D.J.; Brandt, B.W., et al. Interkingdom interactions on the denture surface: Implications for oral hygiene. *Biofilm* **2019**, *1*, 100002, doi:10.1016/j.biofilm.2019.100002.

Brown JL, Johnston W, **Delaney C**, Short B, Butcher MC, Young T, Butcher J, Riggio M, Culshaw S, Ramage G. Polymicrobial oral biofilm models: simplifying the complex. *J Med Microbiol*. 2019 Nov;68(11):1573-1584. doi: 10.1099/jmm.0.001063. Epub 2019 Sep 13. PMID: 31524581.

Delaney C, Kean R, Short B, Tumelty M, McLean W, Nile CJ, et al. Fungi at the Scene of the Crime: Innocent Bystanders or Accomplices in Oral Infections? *Current Clinical Microbiology Reports*. 2018; 5:190-200,doi: 10.1007/s40588-018-0100-3

Kean R, **Delaney C**, Sherry L, Borman A, Johnson EM, Richardson MD, Rautemaa-Richardson R, Williams C, Ramage G. Transcriptome Assembly and Profiling of *Candida auris* Reveals Novel Insights into Biofilm-Mediated Resistance. *mSphere*. 2018 Jul 11;3(4):e00334-18. doi: 10.1128/mSphere.00334-18. PMID: 29997121; PMCID: PMC6041501.

Sztukowska, M.N.; Dutton, L.C.; **Delaney, C.**; Ramsdale, M.; Ramage, G.; Jenkinson, H.F.; Nobbs, A.H.; Lamont, R.J. Community Development between *Porphyromonas gingivalis* and *Candida albicans* Mediated by InlJ and Als3. *mBio* **2018**, *9*, doi:10.1128/mBio.00202-18.

Kean R, **Delaney C**, Rajendran R, Sherry L, Metcalfe R, Thomas R, McLean W, Williams C, Ramage G. Gaining Insights from *Candida* Biofilm Heterogeneity: One Size Does Not Fit All. *J Fungi (Basel)*. 2018 Jan 15;4(1):12. doi: 10.3390/jof4010012. PMID: 29371505; PMCID: PMC5872315.

List of Accompanying Material

Appendices

- I. Supplementary Data: Gene counts from RNA-Sequencing data set of *C. albicans* isolates
- II. Supplementary Data: Metabolomic data set of extracellular metabolomes from *C. albicans* isolates
- III. Significantly enriched pathways from joint pathway analysis
- IV. High resolution versions of figures
- V. Bibliometric analysis on trends in *Candida* biofilm research over the last 20 years

Abbreviations

AAT1: Aspartate aminotransferase
AI-2: Autoinducer-2
AMB: Amphotericin B
ANOVA: Analysis of variance
AS: activity score
CCA: Canonical Correlation Analysis
CFE: Colony forming equivalents
CGD: Candida genome database
CLSM: Confocal laser scanning microscopy
CV: Crystal violet
DEG: Differentially expressed genes
DPD: 4,5-dihydroxy-2,3-pentanedione
DS: Denture Stomatitis
ECM: Extracellular matrix
EDTA: Ethylenediaminetetraacetic acid
FCS: Foetal calf serum
FDR: False detection rate
GC-MS: Gas chromatography–mass spectrometry
GFF: Gene feature format
GO: Gene Ontology
GSEA: Gene set enrichment
GUI: Graphical user interface
HBF: High biofilm formers
HMDS: Hexamethyldisilazane
HOMD: Human Oral Microbiome Database
HPAEC-PAD: High-performance anion-exchange chromatography with pulsed
amperometric detection
HSD: Honestly significant difference
ITS: Internal transcribed spacer
LBF: Low biofilm formers
LC-MS : Liquid chromatography–mass spectrometry
log2FC: log2 fold change
MALDI-TOF: Matrix assisted laser desorption ionization time-of-flight mass spectrometry
MAPK: Mitogen-activated protein kinase
MCA: Multiple correspondence analysis
MCP: Multiple country publication
MS: Mass spectrometry
NCAS: Non-Candida albicans species
NES: Normalised enrichment score
NGS: Next generation sequencing
NMDS: Non-metric multidimensional scaling
NMR: Nuclear magnetic resonance
OHIP: Oral health impact profile

OTU: Operational taxonomic unit
PAPi: Pathway Activity Profiling algorithm
PBS: Phosphate buffered saline
PCA: Principal component analysis
PCoA: Principal coordinate analysis
PD: Periodontal disease
PICRUST: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
PiMP: Polyomics integrated Metabolomics Pipeline
PKA: Protein kinase A
PLS: Partial least squares
PLS-DA: Partial least squares discriminatory analysis
RDP: Ribosomal Database Project
RGR: Relative growth rate
RIN: RNA integrity number
rRNA: Ribosomal RNA
SAB: Sabouraud's dextrose agar
Saps: Secreted aspartyl proteinases
SCP: Single country publication
SEM: Scanning electron microscopy
SGD: Saccharomyces Genome Database
sPLS: Sparse partial least squares
sPLS-DA: Sparse partial least squares discriminant analysis
SRR: Sugar receptor repressor
TCA: Tricarboxylic acid
VIP: Variable importance in the projection
VVC: Vulvovaginal Candidiasis
WOS: Web of Science

1 *Candida* biofilms, their clinical significance and role in interkingdom interactions

1.1 Introduction

Within this thesis introduction we will discuss our current knowledge of *Candida* biofilms. We discuss their clinical implications, factors that drive heterogeneity and differences in phenotype and how they interact with bacteria in the oral cavity. We also discuss the advances in omics and how they are applied to interrogating microbial communities and biofilms. We also discuss our current understanding of oral health, the oral microbiome, and its relationship with *Candida*.

This chapter is also contributed to from work previously published:

Kean R, **Delaney C**, Rajendran R, Sherry L, Metcalfe R, Thomas R, McLean W, Williams C, Ramage G. Gaining Insights from *Candida* Biofilm Heterogeneity: One Size Does Not Fit All. *J Fungi* (Basel). 2018 Jan 15;4(1):12. doi: 10.3390/jof4010012. PMID: 29371505; PMCID: PMC5872315.

Delaney C, Kean R, Short B, Tumelty M, McLean W, Nile CJ, et al. Fungi at the Scene of the Crime: Innocent Bystanders or Accomplices in Oral Infections? *Current Clinical Microbiology Reports*. 2018; 5:190-200,doi: 10.1007/s40588-018-0100-3

Despite their clinical significance and substantial human health burden, *Candida* infections remain relatively under-studied and under-appreciated. The widespread overuse of antibiotics and the increasing requirement for indwelling medical devices provides an opportunistic potential for the overgrowth and colonization of pathogenic *Candida* species on both biological and inert surfaces. In fact, it is now widely appreciated that these are a highly important part of their virulence repertoire. *Candida albicans* is regarded as the primary fungal biofilm forming species and is known for its role in interkingdom biofilm interactions with numerous different bacteria. *C. albicans* biofilms are heterogeneous structures by definition, existing as three-dimensional populations of yeast, pseudo-hyphae and hyphae, embedded within a self-produced extracellular matrix (ECM) (López-Ribot, 2005). Classical molecular approaches driven by extensive studies of laboratory strains and mutants have enhanced our knowledge and understanding of how these complex communities develop, thrive, and induce host mediated damage. Yet clinical observations tell an alternative story, with differential patient responses likely to be a result of inherent

biological heterogeneity from specific clinical isolates associated with their infections. Moreover, the complexity of these infections is increased due to its interaction with host and environmental factors in addition to co-existence with many other bacterial species as interkingdom biofilms.

1.2 *Candida* biofilms

Candida spp. are eukaryotic fungal species of the division Ascomycota that are ubiquitous and exist within the normal flora on mucosal membranes, skin and gut. *Candida* spp. are present in up to ~70% of healthy humans but an overgrowth of *Candida* leads to dysbiosis such as in the case of immunocompromised patients (Perlroth et al., 2007). *Candida* spp., as pathogens, primarily cause candidiasis at the mucosa of the mouth, gut and vagina (Moyes and Naglik, 2011). In immunocompromised patients *Candida* causes invasive candidiasis or candidemia of the bloodstream which is a leading cause of bloodstream infections (BSI) (Rajendran et al., 2016d). *Candida* spp. typically associated with disease include *C. albicans*, *C. glabrata*, *C. tropicalis* and *C. krusei*. *C. albicans* is typically considered the most prevalent fungal species and most dominant fungal species associated with disease in humans (Nobile and Johnson, 2015). Although many *Candida* species cause infections of the mucosa, *C. albicans* is the most commonly recovered in clinical cohorts and this is attributed to its increased pathogenic factors (Nobile and Johnson, 2015). Key to this is its ability to phenotypically switch from yeast to hyphal and pseudo-hyphal forms. Enhanced filamentation compared to other *Candida* species which are only capable of yeast or pseudo-hyphae is thought to explain the higher prevalence of *C. albicans* in mucosal and BSI infections (Rajendran et al., 2016d, Sherry et al., 2014). Hyphal formation provides increased tissue invasion, host evasion and biofilm formation. Throughout this introduction we will discuss the process and implications of hyphal and biofilm formation and how this influences disease and its interaction with the host and other microorganisms.

Microbes rarely exist planktonically as a single entity, instead the majority of organisms are found as complex heterogeneous sessile communities on abiotic and

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

biotic surfaces defined as polymicrobial biofilms (Peters et al., 2012). It has been suggested that up to 80 percent of bacteria favour a co-operative existence (Ramage et al., 2014). It is estimated that two thirds of all infections are due to the presence of a microbial biofilm (Potera, 1999, Sintim and Gursoy, 2016). Biofilms are considered as surface attached or aggregates of microbial cells (bacteria, fungi, and protozoa) that are encased in some form of polymeric substrate. These tenacious and sticky structures are classically recalcitrant to antimicrobial therapies, difficult to remove, and drive pro-inflammatory chronic immune responses that can damage host tissues and materials. *C. albicans* is considered the paradigm for fungal biofilms and significant strides have been made in elucidating how it forms and the mechanistic control of this. *C. albicans* is a normal commensal of human mucosal surfaces and opportunistic pathogen in immunocompromised patients. This dimorphic yeast exists in budding-yeast and hyphal forms, and these morphologies lend themselves to creating a structurally complex biofilm. A simplified model of this begins with yeast cells attaching to a relevant surface via defined adhesins. A microcolony is formed and yeast cells undergo morphological switching to pseudo- and true hyphae, rapidly forming an intertwined meshwork of hyphae interspersed with budding yeast cells. A glucan rich polymeric matrix then encloses the biofilms as it matures, providing protection from host defences and antimicrobial agents. The hypoxic stressful environment within the mature biofilm leads to further filamentation, but also release of yeast daughter cells can then disperse from the site of biofilm formation under physical force, such as flow, and attach to a new substrate and biofilm development starts again. Figure 1.1 illustrates this development cycle and the morphological transitioning. Microscopic analysis has demonstrated that *C. albicans* biofilm formation could be separated into three distinct developmental phases: Early (0 to 11 hours); Intermediate (12 to 30 hours); and Maturation (31 to 72 hours). Confocal laser scanning microscopy (CLSM) has shown mature *C. albicans* biofilms to be complex 3-D structures that can range from anything between 50 to 350 μm thick, depending on the model utilised (Chandra et al., 2001).

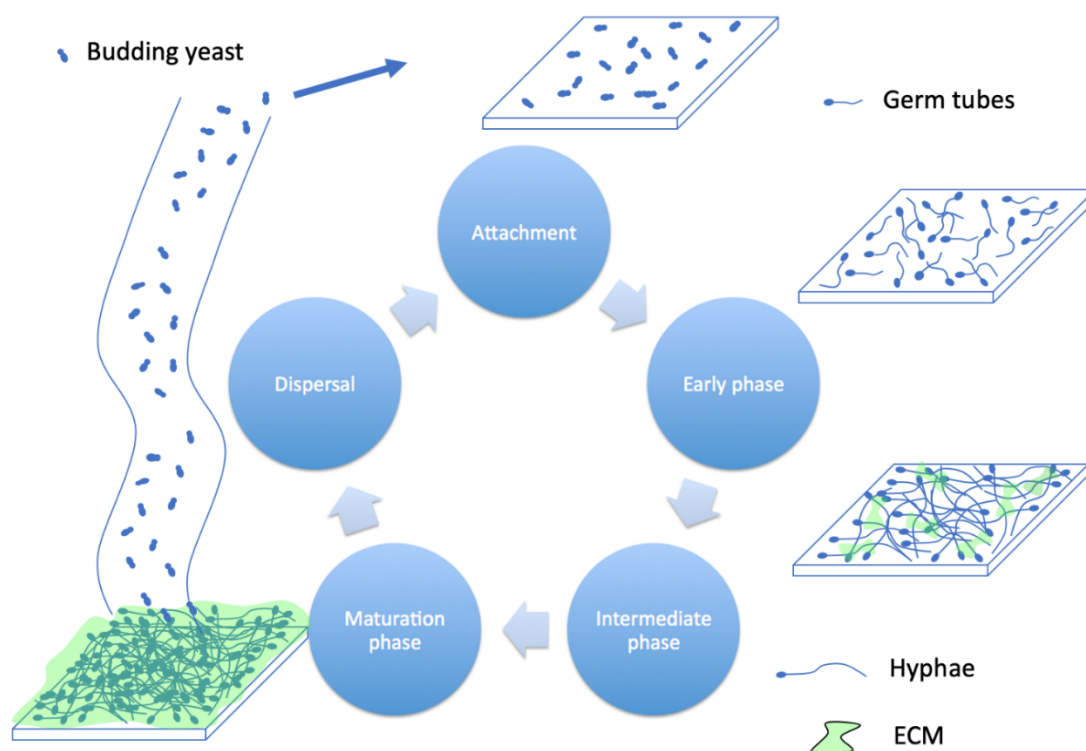


Figure 1.1 **Simplified developmental cycle of *C. albicans* biofilm** (Ramage and Williams, 2012).

1.2.1 Adhesion and virulence

Adhesion of yeast cells is mediated by the agglutinin like sequence proteins ALS1/ALS3 and the GPI anchored cell wall protein Eap1p (Zhao et al., 2006, Li et al., 2007). Once microcolonies are formed, morphological transitioning is governed under the regulatory control of Efg1p (Ramage et al., 2002a). This regulator leads to the colonisation of pseudo- and true hyphae mixed biofilms, additionally hyphal wall protein Hwp2 and putative glycosylphosphatidylinositol (GPI) lipid anchored protein Pga1, Ihd1 and Hyr1 have also been implicated in initial adhesion. Following adhesion the yeast wall protein Ypw1 *S. cerevisiae* homolog is believed to maintain adhesion during early biofilm development (McCall et al., 2019).

Key to successful colonisation and host damage to a mucosal niche is the secretion of various hydrolytic enzymes. These secreted proteins are a primary attribute within the virulence armamentarium of the organism allowing it to invade host tissue, and include proteinases, haemolysin and phospholipase. Of these enzymes, the secreted aspartyl proteinases (Saps) are the most studied, comprising a family of ten genes

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

(*SAP1-10*). The secretion of these enzymes has been attributed with increased virulence, with high levels of expression observed from a variety of diseases including infections of the oral cavity, bloodstream, vagina and patients with diabetes mellitus (Naglik et al., 2008, Joo et al., 2013, Ramage et al., 2012a). Given the diversity of the Sap family, then differential expression of independent genes has been associated with varying anatomical location (Naglik et al., 2003, Joo et al., 2013). During biofilm formation, *SAP5* is up-regulated, significantly correlating with biomass (Sherry et al., 2014). Indeed, an integrated global substrate and proteomics approach identified *SAP5* and *SAP6* as the major biofilm related proteases utilised by *C. albicans*. Manipulation of both genes resulted in decreased adhesion and impaired biofilm development both *in vitro* and *in vivo*, highlighting their role as potential biofilm biomarkers (Winter et al., 2016). Recent studies have identified a novel fungal toxin termed Candidalysin, a hyphae-specific peptide critical for epithelial damage (Moyes et al., 2016) and expression of the gene coding this toxin (*ECE1*) was shown to be highly up-regulated in *C. albicans* isolates capable of forming biofilms (Rajendran et al., 2016b).

1.2.2 Biofilm matrix

Through experimental advancement and use of more sophisticated technologies, *Candida* biofilm ECM has been extensively analysed (Zarnowski et al., 2014, Nett et al., 2010a, Nett et al., 2010b, Mitchell et al., 2016). Compositionally, the ECM is comprised of four main macromolecular constituents: proteins, carbohydrates, lipids and nucleic acid. However, through use of a multi-omics approach, Zarnowski *et al* (2014) identified an abundance of novel components within these subclasses, generating a distinguished compendium of its constituents. This demonstrated its clinical relevance of providing biofilm stability, sequestration of drugs and protection from the surrounding environmental stressors as well as subsequently facilitating biofilm dispersal (Zarnowski et al., 2014). While the majority of ECM-mediated research has focused on the role of polysaccharides, another notable component is extracellular DNA (eDNA) (Martins et al., 2010). Despite only contributing to 5% of the ECM, eDNA plays a substantial role in maintaining structural homeostasis within the matrix. It is thought to act as molecular glue, facilitating cohesion between the other

matrix constituents. Exogenous addition and enzymatic depletion of eDNA have been shown to have both positively and negatively influence biofilm formation respectively (Martins et al., 2010). Furthermore, the addition of DNase to amphotericin B and caspofungin enhances their activity against sessile communities, however no positive interaction is observed with azoles (Martins et al., 2012).

1.2.3 Molecular basis and transcriptional regulation

Given the complexity of the biofilm formation process it is unsurprising that it is determined by a variety of transcriptional regulations. Central to this is the master regulatory transcriptional network as defined by Nobile and colleagues (2012) (Nobile et al., 2012). Originally, a hub of six regulatory genes (TEC1, NDT80, ROB1, BRG1, BCR1, EFG1) was identified that regulate both themselves and approximately 1000 genes involved in biofilm formation processes such as hyphal morphogenesis, ECM production and drug resistance (Nobile et al., 2012). Furthermore, this same group identified an additional three regulatory genes responding to temporal changes in biofilm formation. Using deletion strains, they identified FLO8 as a regulator throughout all stages of development from initial adherence to fully mature biofilms, whereas RFX2 and GAL4 are required only in the later stages of maturation (Fox et al., 2015). While these approaches provide invaluable insights into the transcriptional mechanisms underpinning biofilm development, their limitations lie within only considering laboratory reference strains. Indeed, when the transcriptional profile of a group of *C. albicans* LBF and HBF were compared, no transcriptional differences of two of the master biofilm regulators (BCR1 and EFG1) was shown, despite the phenotypic and biological differences between the strain subsets (Sherry et al., 2014). Figure 1.2 illustrates our current understanding of the molecular basis of *C. albicans* biofilm development.

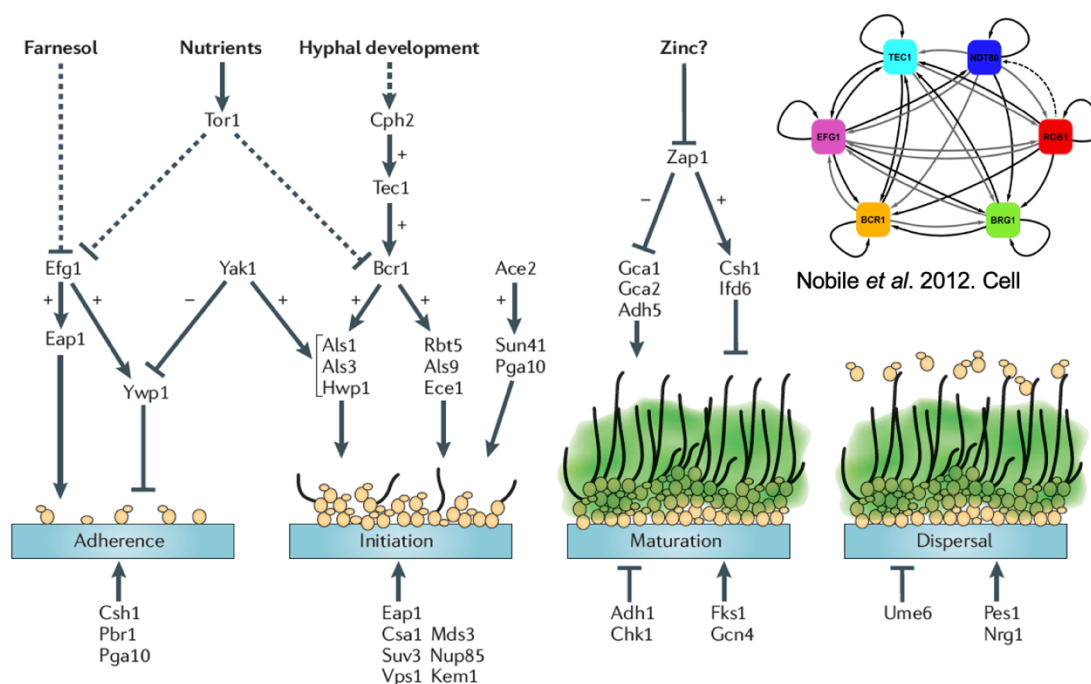


Figure 1.2 **Molecular basis of *C. albicans* biofilm formation** (Finkel and Mitchell, 2011, Nobile et al., 2012).

1.2.4 Metabolism

As well as a defined transcriptional network governing biofilm formation, various metabolic circuits control the transition from planktonic cells to biofilm maturity. Using a metabolomics approach, Zhu and colleagues (2013), performed a time-course analysis of the metabolome of *C. albicans* biofilm development (Zhu et al., 2013). They identified 31 metabolites that were differently expressed between planktonic and biofilm cells that were involved in various processes including the TCA cycle, amino acid biosynthesis and oxidative stress. Interestingly, they showed that trehalose was highly up regulated after 6 hours of maturation. Using a *TPS1* knockout, they demonstrated an impaired biofilm phenotype, as well as increased sensitivity to amphotericin B and miconazole, thus highlighting the importance of the trehalose biosynthesis pathway for biofilm maturation (Zhu et al., 2013). Metabolomics has been used in numerous studies to investigate *Candida* biofilms under different environmental conditions. Such as those that illicit a modified biofilm response. Metabolic reprogramming is an important consideration in many biofilm studies, and we discuss this in further detail in section 1.6.6.

1.3 Biofilm heterogeneity

As highlighted in the previous sections, classical molecular microbiological approaches have shown that deletion or over expression of particular genes enables us to definitively deduce their function in *C. albicans* biofilms. Reinforced by structural biology studies, these tactics have enabled us to deduce the structure/function of specific proteins within the context of biofilm development. Nevertheless, this assumes that molecular manipulations do not have any pleiotropic effects, nor does this take into account inherent biological heterogeneity that bears itself amongst a range of clinical isolates. This begs the question whether using laboratory strains is the optimal way to develop our understanding of microbial pathogenesis (Fux et al., 2005), or instead, whether taking a combinatory approach through evaluating both phenotypic and genotypic characteristics of clinical isolates would enhance our understanding. This section focuses on *C. albicans* biofilm heterogeneity and attempts to examine the literature with respect to what insights can be garnered from working with clinical isolates and observing the inherent heterogeneity that exists. Biofilm heterogeneity for any micro-organism is the difference observed in biofilm formation from a few patchy cells to thick multi layered masses (Wimpenny et al., 2000). Further to this, biofilm heterogeneity in *Candida* biofilms involves cells at different stages of morphology, fluctuating levels of ECM and differences in thickness.

1.3.1 Is heterogeneity clinically important?

Since the earliest descriptions of *Candida* biofilms great strides have been made to unequivocally demonstrate their clinical significance, despite perceived contention in the field. Throughout the human host, *Candida* biofilms colonize a wide variety of anatomical locations, as shown in Figure 1.3. The oral and vaginal epithelium provide a mucosal niche for biofilm formation, whilst indwelling medical devices such as prosthetic heart valves and central venous catheters provide an inert, abiotic substrate for subsequent biofilm adherence and proliferation (Kojic and Darouiche, 2004, Ganguly and Mitchell, 2011). Irrespective of isolation site, biofilm heterogeneity has been reported within the oral cavity, bloodstream, and urinary tract (O'Donnell et al., 2017, Rajendran et al., 2016d, Sherry et al., 2017, Jain et al., 2007, Bitar et al., 2014,

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

Alnuaimi et al., 2013). Within a clinical setting, intravascular catheters provide an optimal environment for *Candida* spp., allowing for the development and maturation of biofilms to which cells can disperse and subsequently cause candidemia. Dispersed biofilm cells have been shown to be more pathogenic than their planktonic counterparts, exhibiting greater cytotoxicity and virulence *in vivo* (Uppuluri et al., 2010). Therefore, the role of the biofilm phenotype has potentially profound implications within the clinical environment. An initial study from Tumbarello and colleagues (2007) aimed to identify the top risk factors associated with mortality rates in candidaemia patients. Using multivariate analysis, they were able to distinguish inadequate antifungal therapy (OR 2.36, $p=0.03$), APACHE III (OR 1.03, $p<0.001$) and overall biofilm-forming *Candida* species (OR 2.33, $p<0.007$) as significant variables associated with mortality (Tumbarello et al., 2007). When scrutinized at the *Candida* species level, only *C. albicans* (OR 3.97, $p<0.001$) and *C. parapsilosis* (OR 4.16, $p=0.03$) were shown to significantly correlate with biofilm-based mortality. A follow up study subsequently identified central venous and urinary catheters, use of total parenteral nutrition, and diabetes mellitus as independent factors that were associated with BSI caused by high biofilm forming isolates (Tumbarello et al., 2012). Furthermore, they demonstrated the potential economic burden of these isolates resulting from increased lengths of hospital stays and use of antifungals and ultimately resulted in an increased possibility of mortality (Tumbarello et al., 2012). A more recent, prospective analysis subsequently identified line removal ($p=0.032$) as a significant risk factor associated with mortality rates from a candidaemia patient cohort, with the removal of an indwelling line correlating with a more positive patient outcome (Rajendran et al., 2016d). Interestingly, when this was subsequently assessed at *Candida* species level, survival analysis demonstrated significantly higher survival rates for patients with *C. albicans* associated line removal compared to no removal, with no differences observed in non-*Candida albicans* species (NCAS) (Rajendran et al., 2016c). Published guidelines have suggested that catheter-related bloodstream infections should result in the direct removal of such devices, if possible (Cornely et al., 2012, Koehler et al., 2014, Mermel et al., 2009). Furthermore, a meta-analysis of seven clinical trials revealed that the removal of central venous catheters significantly correlated with reduced mortality rates (OR 0.50, $p<0.001$) (Andes et al., 2012).

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

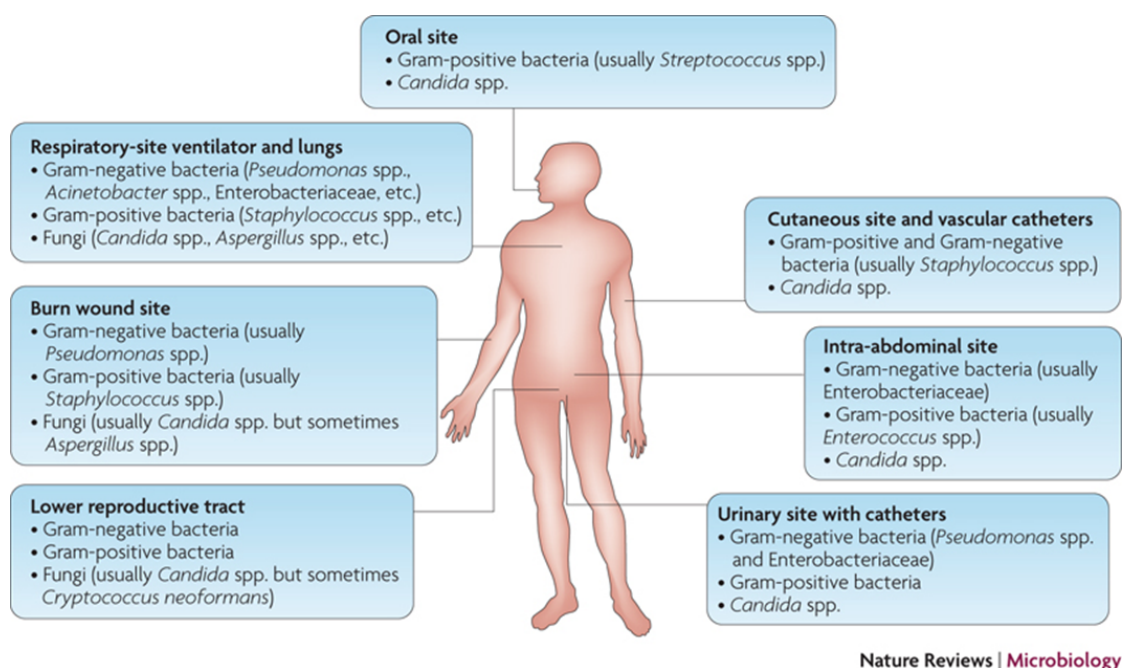


Figure 1.3 **Mucosal and medical devices associated *Candida* biofilm infections** (Peleg et al., 2010).

1.3.2 How does heterogeneity impact antifungal treatment?

Antifungal tolerance is a complex, multifactorial process which can either be induced in response to a compound or manifest as an irreversible genetic alteration as a result of prolonged drug exposure. While resistant planktonic cells predominantly arise from inherited traits to maintain a tolerant phenotype, biofilm resistance rises through mechanisms such as over-expression of target molecules, efflux pump activity and through the protective barrier of the extracellular matrix (ECM) allowing limited diffusion. Undoubtedly the most defining characteristic of biofilms is this intrinsic and adaptive recalcitrance to many antimicrobial therapies. Compared to their free-floating planktonic equivalents, up to 1000-fold higher concentrations of antifungal agents can be required to effectively kill *Candida* biofilms *in vitro*, with the same decreased sensitivities also observed *in vivo* (Ramage et al., 2001b, Kucharikova et al., 2010). Figure 1.4 illustrates the factors associated with antifungal tolerance in biofilms.

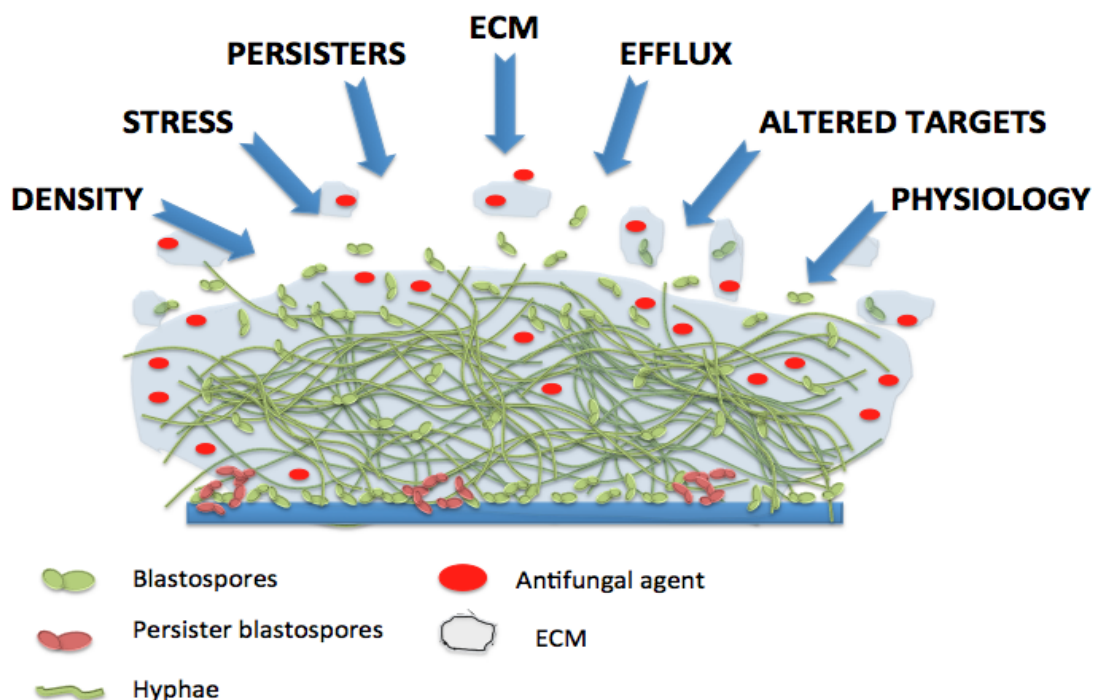


Figure 1.4 **Schematic representation of factors driving biofilm-associated antifungal tolerance** (Ramage et al., 2012b).

Several clinical observations have associated the ability to form biofilms with mortality, but also with azole and inadequate antifungal use. Many studies have sub-categorised *C. albicans* isolates as low biofilm formers (LBF) and high biofilm formers (HBF) (Marcos-Zambrano et al., 2016, Muadcheingka and Tantivitayakul, 2015, Rajendran et al., 2016d). Phenotypically, biofilms formed by these isolates are distinct, with LBF existing predominantly as sparse populations of yeast cells and pseudo-hyphae, whereas HBF have a dense, tenacious hyphal based morphology. *In vivo* there is also biological differences, with increased mortality rates observed in HBF compared to LBF (Hasan et al., 2009, Sherry et al., 2014). Additionally, it was shown *in vitro* that isolates categorised as LBF and HBF were differentially sensitive to azoles and echinocandins at both low and high dosage, with the later less susceptible to these concentrations (Rajendran et al., 2016d).

1.3.3 What drives biofilm heterogeneity?

Interestingly, eDNA is a factor that contributes to the biofilm forming heterogeneity observed between LBF and HBF. Significantly increased quantities of eDNA were

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

released from both early and mature biofilms of HBF strains compared to LBF strains (Rajendran et al., 2014). Given that HBF are more resistant to amphotericin B (AMB) than LBF (Sherry et al., 2014), the combination therapy with AMB and DNase, that sensitises HBF up to 8-fold compared to AMB alone, is very much a matrix-mediated resistance (Rajendran et al., 2014). The role of the other ECM components within biofilm heterogeneity observed in clinical isolates remains unknown, yet given the differences observed between azole and echinocandin susceptibility of these isolates (Rajendran et al., 2016d) it is highly likely that key components are involved and worthy of further scrutiny to determine if strain specific ECM motifs are present.

To better our understanding of the molecular mechanisms facilitating biofilm heterogeneity between *C. albicans* clinical isolates, Rajendran and colleagues (2016) undertook a transcriptional profiling approach (Rajendran et al., 2016b). As expected, well-known biofilm related genes such as *HWP1* and *ALS3* were up regulated in HBF. A non-biased computational approach was further utilized, and in doing so the metabolic circuitry which defined biofilm phenotypes was established (Figure 1.5).

Using KEGG pathway analysis, it was shown that the amino acid pathways arginine and proline metabolism, pyruvate metabolism, and also fatty acid metabolism, were highly upregulated in HBF. Within the subnetwork of these pathways, the gene encoding aspartate aminotransferase (*AAT1*) was shown to be a regulatory hub of these networks. Pharmacological inhibition of this enzyme was shown to perturb biofilm formation, highlighting its potential as a target for biofilm-based infections.

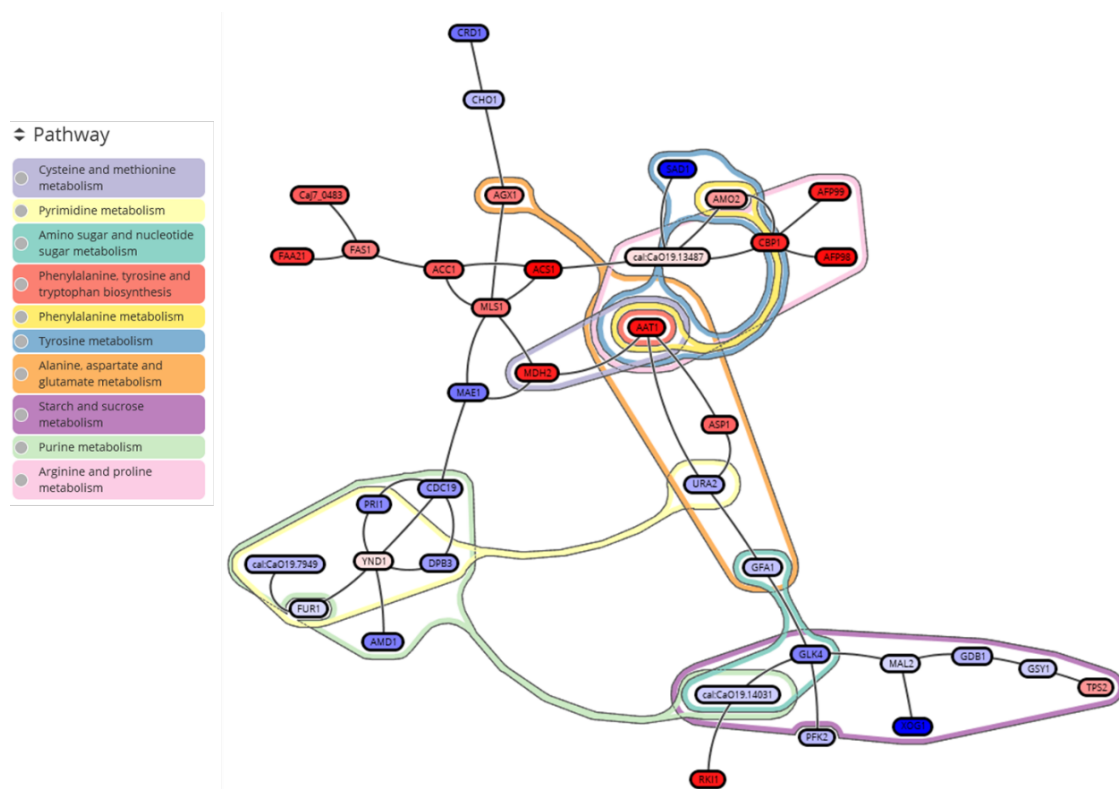


Figure 1.5 **Maximum scoring metabolic subnetwork in the LBF-HBF network.** Differential transcriptional expression between LBF and HBF. Red gene names indicate upregulation in HBF, with blue indicating LBF (Rajendran et al., 2016a).

The adaptation of its metabolism is fundamental to the pathogenicity and survival of *C. albicans* within the host (Brown et al., 2014a). The immune response to *Candida* biofilms is diminished compared to planktonic cells (Nett, 2016), with further evidence suggesting the potential to stimulate biofilm production, resulting in an altered inflammatory output (Chandra et al., 2007). The presence of additional environmental stressors such as pH, thermal and oxidative stress, and also the availability of nutrients results in the metabolic adaptation of the biofilm to acclimatise to its surroundings. Figure 1.6 illustrates an emerging concept, central to this thesis, how different clinical isolates differentially respond to external stimuli. This, combined with inter-relationships with other yeasts and bacteria, creates multiple permutations of strain specific biofilms, all exhibiting their distinct and unique fingerprints.

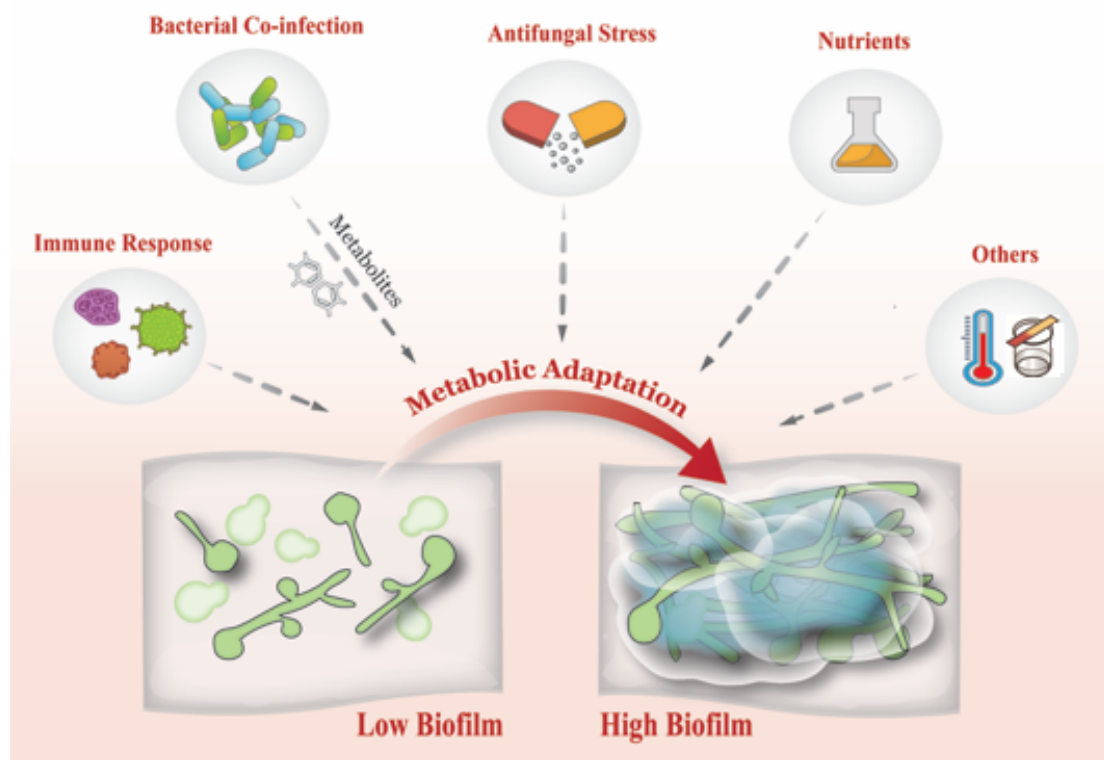


Figure 1.6 **Factors influencing biofilm formation.** There are multiple stimuli that can induce biofilm formation including the immune response, antifungal stress, and bacterial derived metabolites. Environmental stressors can also stimulate biofilm formation, and these include the availability of nutrients, temperature and pH (Kean et al., 2018a).

1.4 Polymicrobial in the oral cavity

Polymicrobial oral biofilms consist of many bacterial and fungal species. Approximately 700 bacterial species or phylotypes (Aas et al., 2005) and more than 100 fungal species have been identified in the oral cavity (Peters et al., 2017). It is estimated that overall species numbers may well exceed 1000, although many of these are uncultivable (Jenkinson, 2011). There is significant diversity in the oral microbiome, varying greatly from individual to individual. For example, only 100-200 species are thought to be found in the oral cavity of any given individual (Paster et al., 2006). Despite this diversity, the concept of microbial “complexes” of microorganisms has emerged, that demonstrates a shift in biofilm colonisation from health to disease, such as in the development of periodontal disease (PD) (Socransky et al., 1998, Haffajee et al., 2008). Our understanding of how dental plaque composition relates to oral health and disease has also changed over time. For example, hypotheses such as

the “non-specific plaque hypothesis” (Theilade, 1986), “specific plaque hypothesis” (Loesche, 1976), “ecological plaque hypothesis” (Marsh, 1994) and “keystone pathogen hypothesis” (Hajishengallis et al., 2012) were all developed over the past 50 years. Over time, these hypotheses have set the foundations of future oral microbiological research, ultimately contributing to our current understanding of the complex nature behind microbial disease onset and progression in the oral cavity. Recently, multi-omics approaches (e.g., genomics, transcriptomics, proteomics, and metabolomics) have enhanced our understanding of microbial interactions in the oral cavity. Theoretically, it is now possible to identify all microbial species that colonise our mouths (Jenkinson, 2011, Dewhirst et al., 2010). The omics platforms provide the power to investigate complex systems in unprecedented detail, and these have been used to examine biofilms in human diseases and in animal models of disease.

1.4.1 *Candida* and the oral cavity

The oral ecosystem is a critical element in oral health, with dysbiosis perceived as the primary driver of disease phenotypes (Marsh and Zaura, 2017). Recent studies have shown that ecological heterogeneity of the salivary ecosystem, its functionality, and its interaction with host-related biochemical salivary parameters are also important considerations in understanding disease processes in the oral cavity (Zaura et al., 2017). However, a notable omission from many of these studies is the importance of pathogenic fungi expressed in terms of the mycobiome, and how these eukaryotes impact the progression of various oral diseases. Indeed, fungi are often neglected in oral microbiology due to their relatively low quantitative contribution within diseased sites of the oral cavity. Nevertheless, consideration of their size may lead to a rethink, as conservative estimates for a yeast cell bio-volume is approximately $70\mu\text{m}^3$ whereas a bacterium is $0.5\mu\text{m}^3$ (Janus et al., 2016). This almost 150-fold difference in the occupancy of available space upon orally relevant surfaces for fungi suggests that it is at least a bystander, but whether a major contributor to pathogenicity or not remains to be determined.

The oral cavity contains numerous different micro-environments, ranging from enamel, mucosa, periodontal pockets, acrylic and metal substrates, and dentine, that

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

are inhabited by tens to hundreds of bacterial species (Xu et al., 2015). Polymicrobial communities can form biofilms upon this vast variety of substrates (Shibli et al., 2008, Diaz et al., 2012, Vieira Colombo et al., 2016). It is now generally well understood that dental plaque biofilms contain many resident species that modulate one another, their environment, and the host response (Millhouse et al., 2014, Hirschfeld et al., 2017, Ramage et al., 2017). Despite the widely acknowledged presence of yeasts within the oral cavity, their active role and participation in oral diseases is generally perceived to be restricted to mucosal-related diseases. Yet, more than 100 fungal species have been reported to colonise the oral cavity and co-exist within complex biofilm populations within aggregates alongside resident bacterial microbiomes (Peters et al., 2017). The emergence of these datasets has been the catalyst for several studies investigating the importance of fungi within oral biofilm infections. This has fuelled the concept of interkingdom communities, which continues to grow above and beyond our traditional viewpoint of bacterial-bacterial interactions (Mukherjee et al., 2017, Bamford et al., 2009, Boisvert and Duncan, 2008). Indeed, the advancement in sequencing technologies has facilitated the characterization of the fungal oral microbiome (Ghannoum et al., 2010, Persoon et al., 2017b, Peters et al., 2017).

1.4.2 Periodontal disease

Periodontitis is a disease, or group of diseases, characterised by a complex host inflammatory response that is stimulated by microbial interactions from complex polymicrobial biofilm plaque. This subsequently leads to damage to the surrounding gingival tissues and supporting structures of the tooth. It varies in severity from reversible gingivitis to severe irreversible periodontitis, where the periodontal ligaments and alveolar bone are destroyed (Page et al., 1978). Despite our ability to easily prevent this disease with oral hygiene measures, it remains one of the most prevalent diseases in the world, with nearly half of adults in the USA developing periodontitis (Eke et al., 2015). Periodontal disease has a well understood bacterial aetiology, with *Porphyromonas gingivalis* considered the keystone pathogen within periodontitis (Hajishengallis et al., 2012). Several other bacteria have also been implicated, including but not limited to, *Tannerella forsythia*, *Aggregatibacter actinomycetemcomitans*, and *Fusobacterium nucleatum* (Megson et al., 2015, Settem

et al., 2012). Given the diversity and numbers of other bacteria in and around the periodontal neighbourhood, then it is difficult to say with absolute certainty that particular periodontal pathogens are the sole cause of periodontitis. Indeed, the concept of oral ecotypes and pathotypes suggests that the sum of different varieties of periodontal pathogens and their functional capacity within the periodontal environment, despite functionally redundant metabolic processes, are more likely to drive synergised virulence leading to clinical disease (Zaura et al., 2017).

Several fungal species have been isolated from the periodontal pockets of patients with periodontitis, with *C. albicans* generally being the most prevalent (Canabarro et al., 2013, Al Mubarak et al., 2013). Notably, the presence of *C. albicans* has also correlated with the severity of periodontitis (Canabarro et al., 2013). Whether it is simply innocently colonizing this environment and playing no active pathogenic role is unknown, yet mounting evidence suggests that it has the capacity to interact with periodontal pathogens and influence their behaviours. Specific bacteria frequently co-isolated with *C. albicans* in periodontal pockets include the anaerobes *F. nucleatum* and *P. gingivalis*. Remarkably, it has been shown that fungi are able to rapidly deplete oxygen within mixed species environments, which may explain why obligate anaerobes and yeasts are observed together (Lambooij et al., 2017). For instance, *P. gingivalis* modulates and enhances the germ tube formation of *C. albicans* (Nair et al., 2001), whereas *F. nucleatum* has been demonstrated to inhibit *C. albicans* hyphal morphogenesis (Bor et al., 2016). Other *P. gingivalis*-related studies however noted an antagonistic effect on the yeast-hyphal transition in *C. albicans*, with *P. gingivalis* notably down-regulating hyphal related genes *ALS3*, *HWP1* and *SAP4* (Cavalcanti et al., 2016b, Thein et al., 2006). Most recently, it has been demonstrated that the attachment of *P. gingivalis* to *C. albicans* is facilitated by the virulence factor InlJ from the internalin protein family, which interacts with the *C. albicans* adhesin *ALS3* (Sztukowska et al., 2018a). Additionally, co-adhesion specific interactions were observed, where adhesive interactions between these pathogens appears to induce the type 9 secretion system of *P. gingivalis*, a system characterised as having an increased community pathogenicity (Hajishengallis and Lamont, 2016).

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

Further studies have shown that enhanced invasion of a gingival epithelial cell line and gingival fibroblasts by *P. gingivalis* is enhanced by pre-incubation with heat-killed cells and the mannoprotein- β -glucan complex from *C. albicans* (Tamai et al., 2011). The mechanism by which *C. albicans* facilitates this invasion is unclear, though the authors hypothesise that the recruitment of *C. albicans* cell wall components increase the recruitment of clathrin in epithelial cells. This is a mechanism by which *P. gingivalis* has been shown to invade host cells (Boisvert and Duncan, 2008). Further physical interactions have been investigated with *F. nucleatum*. This contact-dependent interaction is mediated by the FLO9 *C. albicans* cell wall protein and the RadD *F. nucleatum* membrane protein, which prohibits the morphological switching from yeast to hyphae (Bor et al., 2016). Only recently has this interaction of FLO9 and RadD have been shown to be necessary for co-aggregation of *F. nucleatum* and *C. albicans* under both planktonic and biofilm conditions (Wu et al., 2015). It has additionally been shown that this co-aggregation with *F. nucleatum* has a modulatory effect on the innate immune response. MCP-1 and TNF- α production are reduced during co-aggregation, which Bor et al (2016) conclude has the potential to provide a mutualistic protection from macrophage killing and recruitment of monocytes resulting in an increased persistence (Bor et al., 2016).

Finally, *A. actinomycetemcomitans* has been associated with severe periodontitis (Brusca et al., 2010). *In vitro*, *A. actinomycetemcomitans* adheres to hyphae, although interactions with *C. albicans* show decreased fungal biofilm formation which is mediated by the luxS synthesized autoinducer-2 (AI-2) dependent mechanism. This was also mirrored using the 4,5-dihydroxy-2,3-pentanedione (DPD) synthetic molecule (Bachtiar et al., 2014). In *Streptococcus gordonii*, a ubiquitous oral commensal, it was shown that *C. albicans* hyphal formation was induced in a luxS dependent manner, and that the addition of DPD had no effect on hyphal formation (Bamford et al., 2009). It is evident, as the authors concede, that more studies are necessary to elucidate the interactions between *C. albicans* and *A. actinomycetemcomitans*.

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

There is still much that is unknown and there is conflicting evidence within the literature regarding the importance of interkingdom relationships and their involvement in the progression of periodontitis. The literature does highlight some potentially important synergistic relationships between fungal and bacterial species. However, as is true for many aspects of fungal-bacterial interactions further clinical studies are required to demonstrate functional dependency. A starting point would be to show *in vivo* presence of hyphae within periodontal pockets of patients and associated bacteria.

1.4.3 Denture stomatitis

As the elderly population expands to a predicted two billion by 2050, the number of denture wearers are coincidentally rising. Currently, around 20% of the UK population wear removable dentures of some form, with 70% of UK adults older than 75 years old wearing dentures (Hannah et al., 2017), with many of these individuals suffering from denture stomatitis (DS), an inflammation of the palate (Gendreau and Loewy, 2011). Poor oral hygiene is frequently observed within this patient group and several factors can impact the onset of DS such as salivary flow, denture cleanliness, age of denture, smoking and diet (Martori et al., 2014). Soft tissue inflammation below or above the denture, as a result of persistent exposure to microorganisms, is characteristic of DS (O'Donnell et al., 2017). Microbes frequently adhere to the denture surface and a biofilm quickly develops which can contain numerous species of bacteria and fungi. This is aided by the varied topographical landscape that promotes microbial retention within cracks and crevices of acrylic substrates (Hannah et al., 2017). Denture plaque microbiome studies by our group have identified a variety of oral pathogens, including cariogenic bacteria, such as the *Lactobacillus* species that were positively correlated with high levels of *Candida* spp. (O'Donnell et al., 2015).

Discovery of *C. albicans* and *Lactobacillus* species in denture plaques was unexpected, as these bacteria have previously displayed antagonism with *C. albicans* at other mucosal sites (Parolin et al., 2015). The interactions and mechanisms employed by *C. albicans* and *Lactobacilli* remain somewhat enigmatic. *Lactobacilli* species have

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

demonstrated the ability to inhibit *C. albicans* growth via the release of hydrogen peroxide and fatty acids (Li et al., 2012). Previous to this, in an *ex vivo* experiment, *L. rhamnosus* and *L. reuteri* altered host responses by eliciting an increased inflammatory cytokine response in a *C. albicans* co-infection model (Martinez et al., 2009). Hypothetically, a pro-inflammatory response could exacerbate the inflammation of DS whilst ultimately assisting in the clearance of *C. albicans*. *Lactobacilli* supernatants have been shown to considerably reduce the ability of *C. albicans* to form biofilms (Matsubara et al., 2016). However, the supernatants were unable to significantly reduce the viability of mature biofilms compared to bacterial cell suspensions. It is likely that production of excreted metabolites such as hydrogen peroxide and short chain fatty acids may interfere with initial adhesion, but direct bacterial-fungal interactions occur to disperse mature *C. albicans* biofilms. Patients with more severe DS are colonised with greater numbers of *C. albicans* and *Lactobacilli*, indicating the possibility that these organisms can detect changes in their environment and alter their behaviours appropriately to effectively colonise the oral cavity (Bilhan et al., 2008).

Although *Lactobacillus* species are the most commonly isolated bacteria from DS biofilms, other *Candida* species, namely *C. glabrata*, have also been detected (Coco et al., 2008a). Co-infection with *C. glabrata* results in upregulation of key virulence genes (*ALS3* and *HWP1*) in *C. albicans* (Alves et al., 2014). This increased virulence, in return, complements the ability of *C. glabrata* to invade epithelial tissue (Silva et al., 2011). Authors hypothesised that during penetration of tissues by *C. albicans*, *C. glabrata* is transported into host cells via forming aggregates on the hyphae, as has been shown with various bacteria.

As studies get closer to understanding the mechanisms of interactions within dual-species biofilms, this will pave the way to elucidate how interactions within multispecies denture plaque contribute to disease processes. Indeed, it has been shown *in vitro* that multi-species interkingdom interactions synergise one another, with hyphae induced by *Streptococcus oralis*, and the overall biovolume of denture biofilms further enhanced with accompanying *Actinomyces oris* (Cavalcanti et al.,

2016a). Understanding the behaviours of multi-species biofilms *in vivo* will prove more useful in the management of DS, where rough surface topography plays an additional physical role in supporting detrimental interactions. Though, this can be mitigated by physically altering denture surfaces to prevent *Candida* adhesion (Alalwan et al., 2018).

These sections have highlighted how some *Candida* is a key human pathogen with high variability in its ability to form biofilms, additionally through different interactions it is able to modulate its morphology and virulence. We have discussed how *C. albicans*, as in other host sites such as the vagina and bloodstream, forms a critical role in oral diseases. It undoubtedly plays an important accessory role to a variety of bacterial pathogens. Within the myriad of different microbial interactions that occur within the oral microbiome, there are synergisms and antagonisms, with the most dominant manifesting themselves in disease outcomes. *C. albicans* can physically, metabolically and through the release of soluble molecules, play an important participatory role in the oral diseases outlined above.

1.5 Methodological approaches to understanding *Candida* and mixed-species biofilms

1.5.1 The conventional and established approaches

When screening large collections of clinical isolates from different patient cohorts, several experimental strategies have been utilized, predominantly quantifying biomass using dry weight, stains such as crystal violet, and the metabolic dye XTT (Azeredo et al., 2017). Each technique has their own benefits and caveats, but caution must be taken when interpreting the data achieved from each assay, particularly when correlating it to clinical outcomes. Due to the heterogeneity found between strains and the varying laboratory models and techniques, standardization becomes problematic. One of the most used bioassays is the sodium salt XTT [2,3-bis(2-methoxy-4-nitro-5-sulfo-phenyl)-2Htetrazolium-5-carboxanilide] (Hawser, 1996, Hawser et al., 1998). This biofilm assay is highly reproducible and allows for a high throughput of multiple microtiter plates without compromising accuracy. Its

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

usefulness comes with susceptibility testing, allowing for the direct comparison of antifungal treated samples compared to an untreated control (Ramage et al., 2001a). Given the metabolic variation observed between both different strains and species then caution must be taken when interpreting the assay, as a measurement for biofilm development may simply reflect high cell numbers (Kuhn et al., 2003, Taff et al., 2012). For example, scant biofilms of non-*albicans* yeasts may show a high XTT value, yet minimal biomass is present. Therefore, the output achieved from XTT is only cellular viability and it does not take into account other biofilm components such as the ECM, which are arguably the most important when it comes to biofilms (Nett et al., 2007).

Another commonly used assay for biofilm formation is crystal violet staining. This method provides the total quantification of the biofilm biomass (cells and ECM) and allows for rapid high throughput processing of multiple samples. However, variability of the washing step can result in both over- and under-estimation of biomass, with the assay also unable to differentiate subtle differences between samples (Azeredo et al., 2017). An interesting example of this was described in a recent study, where these techniques were used to stratify the ability of *Candida* bloodstream isolates to form biofilms (Pongracz et al., 2016). There was no clear standard for their stratification to denote strains as biofilm or non-biofilm formers, with crystal violet values of OD₅₇₀ >0.09 simply denoted as a biofilm former. By doing so, it was concluded that NCAS form greater biofilms than *C. albicans*, and that biofilm formation does not correlate to clinical outcomes. This is in contrary to a wealth of previous literature, whereby the ability of *Candida* isolates to form a biofilm does associate with mortality (Rajendran et al., 2016d, Soldini et al., 2017, Tumbarello et al., 2012, Tumbarello et al., 2007).

Discrepancies between these findings illustrates the necessity for standardised testing to elucidate biofilm-related risk factors. The Ramage group have taken a 'belt and braces' approach, using a combinational approach of crystal violet, XTT and SYTO9 fluorescence quantitative biofilm assays. Here, significant correlations were observed in *C. albicans* biofilm formation, which was subsequently used to stratify biofilm-forming ability (Rajendran et al., 2016d). Irrespective of the quantitative approach, wide-spread biofilm heterogeneity is observed within different clinical panels of

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

isolates (Kumar and Menon, 2006, Pongracz et al., 2016, Rajendran et al., 2016d, Sherry et al., 2014). Collectively, these data suggest that different *Candida* strains function differently, and that consideration should be given to the individual isolates as we try and understand their clinical importance with respect to antifungal resistance and pathogenic potential.

Conventional culturing by process of nutrient plates is still a highly effective and efficient way of determining *Candida* characteristics. First stage identification via CHROMagar is a fast and effective way to distinguish between *Candida* spp. It is routinely used in first pass screening of clinical isolates from which improved identification can be performed (Ozcan et al., 2010). Matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF) has become an integral tool in the microbiologist's toolkit for the fast and accurate identification of clinical yeasts. This technique offers an improved method of identification with a higher resolution, capable of identifying more *Candida* spp. than previous techniques (Alizadeh et al., 2017, Yaman et al., 2012).

Microscopy, both light and fluorescence, have remained consistently used in *Candida* biofilm research. Useful in distinguishing between both yeast and hyphal forms and additionally visualising interactions with host and bacterial organisms. One of many examples of the utilisation of light and fluorescent microscopy has been to identify the level of *C. albicans* filamentation, and additionally the ability of *C. albicans* survival strategies in response to macrophages (Uwamahoro et al., 2014). Moreover, microscopy is useful in determining the level of interaction with synergistic and antagonistic bacteria. Methodologies that profile both *Streptococcus* and *Porphyromonas* spp. have been developed, aided by light microscopy to quantify the level of adherence between yeast and hyphal cells of *C. albicans*. Using bacterial and fungal specific probes *C. albicans* cells can be classified as adhered or non-adhered to bacteria allowing for quantification through high levels of repeated fluorescent images (Bamford et al., 2009, Sztukowska et al., 2018b). Additionally, confocal and scanning electron microscopy (SEM) are routinely utilised to identify structural and architectural changes in *Candida* biofilms. Confocal Scanning Laser Microscopy

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

(CSLM), as discussed previously, allows for the identification of overall structure and features such as the thickness of biofilms. It is also useful in biofilm models for determining changes in thickness and architecture of the biofilm in the presence of treatments or changes induced through genetic knockouts (Nobile and Mitchell, 2005). Similarly, those biofilm changes that are elicited by the presence of bacteria can also be visualised using CSLM (Fox et al., 2014).

1.5.2 The new and innovative approaches

Over the past 20 years, the use of high throughput technologies has started to become routine to support biological discoveries and drive hypotheses. Omics approaches are a collection of technologies and bioinformatics approaches that allow for increasingly broad and in-depth analysis of the microbiome, transcriptome, metabolome, proteome, and genome. They can give a holistic view of the genes, mRNA expression, metabolites, and proteins of cell (Horgan and Kenny, 2011, Zhang et al., 2010). Such techniques are now increasingly being used to answer questions regarding interactions and functional properties of microbial communities. The literature shows that there are several clinically relevant sites in which poly-microbial communities are known to exist, comprising of numerous fungi and bacterial species (Harriott and Noverr, 2011, Peters et al., 2012, Sherry et al., 2016b, Thein et al., 2006). Relevant sites include diabetic foot wounds, sinus infections, oral cavity and respiratory tract infections, as well as medical device related infections (Peters et al., 2012). At these sites, fungal-bacterial biofilms, particularly *Candida* species biofilms, are responsible for the occurrence of disease driven by synergistic microbial interactions (Harriott and Noverr, 2011). Omics allow for new avenues of research in understanding these polymicrobial biofilm communities.

1.5.3 Microbiome

The “microbiome” is a term that has been used to mean both all the microorganisms and all the genetic material of microbes in a community. For the purposes of the work in this thesis we are commonly referring to the genomes of all the microbes (Kho and Lal, 2018). Over the last 10-15 years several NGS platforms have emerged including Roche 454, Ion torrent, Nanopore and Illumina based platforms. Due to the cost

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

effectiveness and the high-quality reads provided by the Illumina platform it has become the primary platform for microbiome sequencing (Bharti and Grimm, 2019). The Illumina platforms sequence one base at a time through fluorescence signals in a process named sequencing by synthesis. Fragments of DNA are hybridised to a micro fluid flow cell with use of ligated adapters. Following cluster formation through bridge amplification nucleotides are then introduced to the flow cell each with their own fluorescent marker. Sequential images of the flow cell are taken at the incorporation of the nucleotide into the growing strand and the images are analysed with the colour of each base being recorded. Illumina has been improving their machines generation upon generation, each time providing larger throughput and improved error reduction with the Illumina Miseq/Hiseq becoming the most popular amplicon sequencing platforms. Work within this thesis was carried out using the Illumina MiSeq platform using short reads, however new sequencing platforms such as the MinION by Oxford Nanopore in the future may allow for faster identification of microbiomes on the lab bench. The ultra-long reads allow for the entire amplification of the 16S RNA region (Nygaard et al., 2020). Promising results from these studies which offer a potentially higher resolution and a faster and convenient method for microbiome sequencing.

16S ribosomal RNA (rRNA) gene, the amplicon most used in identifying different bacterial species, is a group of nine regions (V1-9) that are “hypervariable” and present within all bacteria (Chakravorty et al. 2007). Each hypervariable region shows differing levels of consistency in identifying specific groups of bacterial species; it is thought that identification on one region alone can be biased due to “PCR bias and low taxonomic precision” (Guo et al. 2013). The sequences that are produced from the 16S rRNA sequencing are grouped into OTU’s based upon their similarity. Commonly a threshold of 97% is used, which assumes that bacterial strains have this similarity within the 16S rRNA gene and all sequences clustered to this OTU are identified as the same genus or species (Konstantinidis and Tiedje 2005). The accuracy of the 97% similarity threshold however is still under debate. To date one of the biggest limitations in 16S RNA sequencing is the ability to predict taxonomic profiles at the species level (Gao et al. 2017). OTUs are identified against comprehensive

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

taxonomic databases which include the ribosomal database Project (RDP), SILVA and Greengenes (Cole et al., 2014, Pruesse et al., 2007, DeSantis et al., 2006).

In place of the commonly used OTU, similarity filtering based upon a fixed similarity amplicon sequence variant (ASV) is being advocated. ASV methods include the DADA2, Deblur and UNOISE (Callahan et al., 2016, Amir et al., 2017, Edgar, 2016). These methods aim to denoise the data identifying true biological sequence from noise within the data. With biological sequence variants being considered more likely to be repeatedly observed than errors (Callahan et al., 2017).

Due to the cost, ease of analysis and speed of targeted gene or amplicon sequencing has led to its relatively popularity. Alternatively, entire metagenomic screening by shotgun sequencing is an alternative method of microbiome profiling. Metagenomics is comprehensive identification of the entire genetic composition of an environment (Bharti and Grimm, 2019). Metagenomics, despite its cost and increased difficulty, overcomes many of the limitations of 16S amplicon sequencing. Biases are introduced by regions and there are difficulties discerning different species from the 16S alone. Additionally, this method offers improved identification of novel species and also inference of the functional profile of the microbiome (Baker et al., 2019, Al-Hebshi et al., 2019). Functional properties as we have described such as regarding ecotypes have become interesting focuses within bacterial ecology. Tools such as Phylogenetic Investigation of Communities by Reconstruction of Unobserved States PICRUST which infer KEGG functionality from the amplicon target attempt to overcome this limitation (Zhang et al., 2019, Langille et al., 2013). However, functional predictions from the total genetic composition are obviously more desirable.

The oral cavity contains one of the most diverse microbiomes of the human body, second only to that of the gastrointestinal tract (Human Microbiome Project, 2012). Around 280 bacterial species from the oral cavity have been cultured and classified (Dewhirst et al., 2010). However, this is less than half of the 700 species predicted using the Human Oral Microbiome Database (HOMD) and additionally 100 fungal species (Ghannoum et al., 2010, Chen and Dewhirst, 2013). In 2001 the term

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

microbiome was coined by Joshua Lederberg “to signify the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space” (Lederberg and McCray, 2001). The term has since been adopted by investigators of the poly-microbial communities in oral health and disease. Investigation of the microbiome has become an important area of study. This is due to our understanding that oral diseases such as caries, periodontitis and denture stomatitis are related to dysregulation in a wide range of organisms as opposed to only one pathogen (Struzycka, 2014, Abusleme et al., 2013, O'Donnell et al., 2015b).

Microbiome studies in oral disease have demonstrated the existence of distinct microbial communities compared to health. In periodontitis microbiome studies, in contrast to other clinical host sites (e.g., gastrointestinal tract), an increased bacterial diversity and richness in disease has been reported compared to health (Abusleme et al., 2013, Sousa et al., 2017). Additionally, these studies have elucidated previously unappreciated species. Next generation sequencing (NGS) technologies have revealed a greater complexity within the periodontal disease microbiome, well beyond the traditional “red complex” that is comprised of *Tannerella forsythia*, *Porphyromonas gingivalis* and *Treponema denticola* (Rocas et al., 2001). Microbiome studies have shifted this dogma, showing the prevalence of many disease associated genera and species, including but not limited to, *Spirochetes*, *Filifactor* and *Fusobacterium*, highlighting a more diverse disease community than previously considered (Griffen et al., 2012, Chen et al., 2018).

Dysbiosis and the progression from a healthy to disease state is generally considered to be due to a decrease in the diversity of the microbiome (de Paiva et al., 2016, Giloteaux et al., 2016). However, within the oral cavity this assumption is not so straight forward. Within the oral cavity it has been shown that lower levels of diversity and microbial richness is associated with health in relation to periodontal disease, according to the *Shannon* alpha diversity measure (Griffen et al., 2012, Abusleme et al., 2013). Conversely, using multiple alpha diversity it was demonstrated that there was no marked difference in the species richness between periodontal disease and healthy microbiomes (Galimanas et al., 2014). The 2014 study highlights that

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

differences in the findings can be explained by sampling location and methodologies in DNA extraction. Indeed, within the oral microbiome there are diverse ecological niches, within which the microbial composition can vary greatly (Xu et al., 2015). This variation is in part due to the numerous surfaces within the oral cavity which include but is not limited to the hard and soft tissues of the teeth, and between teeth and the soft and hard palate (Aas et al., 2005). The heterogeneity of the oral microbiome increases the challenges of studying it, this is added to by the number of distinct microbiome niches within the oral cavity (Avila et al., 2009).

1.5.4 The mycobiome

In addition to bacterial species, fungal species also play pivotal roles in the oral microbiota of healthy and diseased individuals (Ghannoum et al., 2010). There are approximately 2 million fungal species and of these, 600 are thought to be able to cause infections in humans (Roilides, 2016). Analogously with the identification of the prokaryote microbiome the eukaryotic mycobiome has been defined primarily using amplicon sequencing (Cui et al., 2013). The amplicon utilities is an internal transcribed spacer region (ITS) which encompasses the locus of the 18S, 5.8S and 28S ribosomal subunits genes (Dollive et al., 2012). The eukaryotic component has been termed either the fungal microbiome or the mycobiome (Ghannoum et al., 2010).

Comparatively very few studies have been carried out on the fungal microbiome compared to the bacterial microbiome. However the importance of studying the fungal component and its relationship with the bacterial microbiome has been noted not just in relation to oral health but in regards to skin, lung and GI tract microbiomes (Huffnagle and Noverr, 2013). Oral mycobiome studies to date have sought to validate the protocols and build a baseline of the healthy community within the oral cavity (Dupuy et al., 2014, Ghannoum et al., 2010).

Within the gut of Crohn's disease patients, a form of Inflammatory Bowel Disease (IBD), there are shifts in both the fungal and bacterial microbiome with some bacterial and fungal species being correlated. For example, *Saccharomyces* species are negatively associated with most bacteria whilst *C. tropicalis* is positively associated

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

with pathogenic bacteria (Hoarau et al., 2016). Another study regarding IBD showed changes in fungal biodiversity, with significant changes in eukaryotic species such as *S. cerevisiae* being associated with disease. *Saccharomyces* and *Malassezia* in this study were shown to be positively associated with numerous bacterial species (Sokol et al., 2017). These studies are indicative of inter- and intra-kingdom interactions that can occur in the transition from a healthy to disease state.

In 2010, a study led the way in characterising the fungal component of the oral microbiota via amplification of the internal transcribed spacer (ITS). Unlike 16S sequences with fixed length amplicons, ITS sequences can produce variable sequence lengths, which makes bioinformatic processing more challenging. The authors observed that the most common genera of fungi were *Cladosporium*, *Aureobasidium* and *Saccharomycetales* (Ghannoum et al., 2010), many of which were subsequently confirmed by Dupuy and colleagues (Dupuy et al., 2014). Interestingly, amongst the observed operational taxonomic units (OTUs) were several genera that were not in consensus with the initial study, including *Saccharomyces* and taxa including the *Saccharomycetales* order that were not found in as high frequency as reported initially (Ghannoum et al., 2010).

In more recent studies, comparisons between oral health and disease have been made. For instance, a study comparing the mycobiome of periodontal disease and healthy individuals reported similar levels of fungal species (over 100) compared to the first report (Ghannoum et al., 2010). However, in their cohort it was observed that *Candida* and *Aspergillus* were the most frequent genera, being present in 100% of samples. Interestingly, there was no significant difference in the overall diversity of fungal taxa between periodontal disease and the healthy cohort, or the overall composition. An increase in abundance the *Candida* genus in periodontal disease compared to health was observed, although this was not found to be statistically significant (Peters et al., 2017). Notably, this was a pilot study and was perhaps hampered by its limited cohort size. Future mycobiome studies with larger cohorts and more stringent classification of disease state and other metadata would provide a more comprehensive interpretation of the mycobiome in oral health.

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

With 16S now being well established, and ITS amplicon sequencing becoming more readily available, there is now the ability to perform co-occurrence studies (Mukherjee et al., 2017, Persoon et al., 2017a). Dual bacteriome and mycobiome analysis of sites within oral cavity (endodontic) infections has been demonstrated by Persoon and colleagues (2017). This group demonstrated a co-occurrence of acidogenic bacteria in the presence of fungal species, and an overall positive correlation of *C. albicans* with bacterial species. Negative correlations of bacterial species were observed with increased *C. dubliniensis*. These authors highlight the current limitations of these approaches, including increased difficulty in DNA extraction, PCR amplification due to length variation of ITS, and inconsistent fungal nomenclature. The oral mucosal mycobiome and microbiome were assessed regarding oral lichen planus in 2019. They were able to discern correlations between fungal and bacterial species in both diseased and healthy cohorts. This study found higher levels of the genera *Candida* and *Aspergillus* in patients with oral lichen planus compared to healthy. Correlating with the increase abundance of *Candida* in disease was the increased abundance of pathogenic aerobes. The authors additionally note how the mycobiome is often overlooked in studies of this type and studies such as this highlight the potential benefits of profiling both the mycobiome and bacteriome (Li et al., 2019). Despite these issues, as these types of studies grow and analytical pipelines become more developed, then the possibilities for understanding complex interkingdom interactions will become more fully realised.

1.5.5 Transcriptome

RNA-Seq or whole transcriptome shotgun sequencing and related processes allow for the whole of the gene expression data to be analysed at a given point in a time. RNA is isolated from the micro-organisms of interest and the mRNA is selected and then synthesised to cDNA, amplified and sequenced (Tan et al., 2015). Sequencing is typically through Illumina sequencing as previously described. Sequenced data undergoes computational processes or a pipeline of analysis which typically involves quality control and clean-up of the reads, alignment, mapping to the genome, quantification, and differential expression analysis of the genes (Conesa et al., 2016a). In contrast to genomic studies, this actual expression compared to potential

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

expression can lead to more accurate functional understanding of micro-organisms in various conditions (Horgan and Kenny, 2011). In microbiology, and specifically biofilm research, transcriptomics has allowed for the regulation of pathways to be shown such as fat metabolism, regulation of DNA repair genes and difference in regulation at different dynamic states (Nakamura et al., 2016). Within oral samples analysis of the transcriptome has been utilised to identify signatures of dysbiosis. Using databases of known and predicted functions, functional inferences can be made of the expression data. Libraries such as the KEGG and GO, which are compilations of what is currently known about a gene including its associated processes, importance in resistance, metabolism processes, signalling pathways and other key processes can be identified (Nakamura et al., 2016, Li et al., 2015). Metatranscriptomics is able to build upon metagenome/microbiome studies allowing for further inference of the functional activity of dysbiotic biofilms. For example in 2018 an oral study correlated the fluctuations in community composition with the functional changes in malate and lactate dehydrogenase and novel pathways in response to decreasing pH (Edlund et al., 2018).

Previously, Rajendran and colleagues (2016) used profiling of the transcriptome to define *C. albicans* biofilm heterogeneity, demonstrating key metabolic pathways driving the biofilm phenotype (Rajendran et al., 2016b). A study which is a precursor for the work outlined within this thesis. The authors demonstrated the effectiveness of profiling gene expression to discern differences in low and high biofilm forming *C. albicans*. Transcriptomics has been utilised in interrogating *in vitro* biofilm models and for monitoring biofilm development from initial to mature stage in the biofilm cycle. Important information can be gleaned in biofilm models of many organisms including *C. albicans*. These include changes in lifestyle, biofilm-host, microbe-microbe interactions, and biofilm resistance. Biofilm profiling through transcriptomics has been previously shown in bacterial species *Porphyromonas*, *Streptococcus*, *Pseudomonas*, *Enterococcus* and *Candida* species (Romero-Lastra et al., 2019, Wu et al., 2019, Partoazar et al., 2019, Cheng et al., 2019, Kean et al., 2018b). Commonly transcriptomics by RNA-Seq has been used to compare planktonic to biofilm counterparts and biofilm dynamics in response to antimicrobial treatments.

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

Identification of biofilm related drug transporters, which were linked to drug resistance, in the multi-drug resistant emerging pathogen were discerned using transcriptomics. Transcriptome assembly and profiling of *C. auris* was instrumental in identifying key resistance features in this clinically important pathogen. Highlighting differences in efflux pump expression related to resistance (Kean et al., 2018b). Transcriptome analysis of mixed species biofilms has also emerged as a useful way of profiling bacterial-bacterial and fungal-bacterial interactions. Interkingdom interactions from both fungi and bacteria are possible to discern through a more complex method of ribosomal RNA depletion. Read alignment also requires reads first being aligned to the fungal species before then remaining unaligned reads being aligned to the bacterial species. This methodology has been successfully applied in *C. albicans-Streptococcus* and *C. albicans-P. gingivalis* interactions (Dutton et al., 2016, Sztukowska et al., 2018b, Ellepola et al., 2019).

1.5.6 Metabolome

Metabolomics provides a snapshot view of the whole metabolite profile of a cell at a given moment under specific conditions. Quantification of the metabolome is performed either by mass spectrometry (MS) or nuclear magnetic resonance (NMR) (Schauer et al., 2005, Villas-Boas et al., 2005, Markley et al., 2016). The use of metabolomics in microbiology and specifically biofilm research allows for the identification of pathways within target organisms. This pathway analysis informs the overall aetiology of single or polymicrobial communities (Washio et al., 2010). The use of novel mass spectrometry techniques has been demonstrated to be a powerful tool in the investigations of fungi-bacterial interactions in a biofilm model (Weidt et al., 2016). A study in which the associations between *Candida* and *S. aureus* were inferred through the detection of intra and extra-cellular metabolites.

In contrast to more typical biological methods, untargeted metabolomics allows for a whole and unbiased view of the chemical processes (Weckwerth, 2010). Metabolites can be indicators of pathogenicity and upstream pathways or directly responsible for the pathogenicity of an organism (Bien et al., 2011, Stipetic et al., 2016). Using metabolites as a method of discerning a state of illness or biological system is an old

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

concept in medical biology (Tang, 2011). The progress of MS and NMR technologies allow for this concept to be integrated with other omics and has become an important component of systems biology. Complementary to untargeted metabolomics targeted metabolomics which aims to quantify with greater sensitivity a select number of metabolites. These metabolites are discerned by standards, usually a pre-defined group of high-quality core compounds (Weidt et al., 2016). Either targeted or untargeted metabolic fingerprinting is the process of determining the intracellular metabolic activity and metabolic footprinting infers the metabolic profile from the supernatants (Aggio et al., 2010). Both footprinting and fingerprinting are commonly performed on microbiological cultures using the analytical method of mass spectrometry. This process is often preceded by chromatography, such as liquid chromatography; the process of passing metabolites through liquid within a column. An example is the pHILIC columns, used for separation of the analytes as they interact with liquid phase. The time to pass through the column and reach the detector then determines the retention time (RT) for the analytes (Xiang et al., 2003). Due to the vast number of metabolites in microbial cultures, analytes must first be separated before MS detection by machinery such as the Orbitrap Fourier transform analyser (Hu et al., 2005). The Orbitrap traps ions in orbit around a spindle via quadrupole and converts the image of the trapped ions in a mass spectrum m/z by fourier transform. LC-MS by this and similar methods can then be applied to identify the majority of biological metabolites. The RT and mass can then either be used in global untargeted fashion to identify metabolites against databases containing known masses or compared to pure standards for accurate RT and mass analysis resulting in a more targeted analysis.

Targeted/untargeted metabolomics allows for a complementary analysis method leading to the identification of expected compounds and additionally more exploratory analysis. For example, in discerning the differences in mixed *C. albicans*-*S. aureus* biofilms which was demonstrated using targeted/untargeted metabolomics (Weidt et al., 2016). From spent media they were able to determine sugars, amino acids and organic acids that were differentially spent or excreted in the metabolic footprint of *C. albicans* single or *C. albicans*-*S. aureus* dual species biofilms. This

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

metabolomic footprinting allowed for inferences to be made about the modulation of the pentose phosphate pathway in dual species biofilms compared to single species. Untargeted fingerprinting analysis of *C. albicans* biofilms was also capable of discerning metabolomic reprogramming of amino acid, ergosterol, lipid and sugar metabolism in response to hypoxia. The authors utilised ultrahigh-performance liquid chromatography–tandem mass spectrometry on *Candida* lab strains grown at 10 to 180 minute timepoints in the presence and absence of hypoxic conditions. Hypoxia was induced in biofilms grown in media by reducing the atmospheric conditions to 5% oxygen. This study highlights the global metabolomic shifts that are associated with shifts in biofilm formation under different environments which could only be elucidated through high throughput methods (Burgain et al., 2020). Within this study there is complex rerouting of metabolic pathways such as carbon metabolism, pentose phosphate pathway and amino acid metabolism, that is undertaken by polymorphic fungi in response to stress and environment. They additionally were able to correlate their analysis to the transcriptome of the same samples. Similarly, gas chromatography-mass spectrometry has been utilised to identify changes in metabolism caused by gene knockouts in *C. albicans*. Metabolomic fingerprinting was able to identify metabolites that differed in their intracellular metabolic ratio in glutamate dehydrogenase mutants compared to wildtype in hyphae inducing conditions (Han et al., 2019b). They propose that under environmental stress *C. albicans* will globally downregulate central carbon metabolism. They also proposed that under stress *C. albicans* was unable to assimilate nitrogen sources from the proline rich media filamentation due to the lack of glutamate dehydrogenase. They therefore propose nitrogen metabolic pathways as targets for therapeutic potential in inhibiting hyphal morphogenesis.

In addition to mass spectrometry, a fragment spectrum of metabolites can be utilised to distinguish between analytes with same m/z value for improved resolution of untargeted metabolomics. Fragmentation patterns of peaks can be compared to known spectrums to infer which analyte is present with the sample. This has also been used to infer greater detail from untargeted metabolomics in biofilm and in particular *Candida* biofilm hypoxia studies previously mentioned (Burgain et al., 2020).

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

Global metabolomics studies provide snapshots at a given point in time, they do not give much information of directionality through metabolic pathways or rates. Metabolic flux analysis aims to overcome these limitations using radioisotopes. Metabolic flux of these isotopes can be monitored and tracked as they are utilised and shuttled through metabolic pathways within the target cells, also highlighting changes in metabolic turnover rather than concentration alone. This has been demonstrated in a number of biofilm models in *P. aeruginosa* comparing fluctuations in pathways between planktonic and biofilm models (Wan et al., 2018). Similarly, in *C. albicans*, isotopes were utilised to identify changes in metabolomics of the glutamate dehydrogenase *gdh2* and *gdh3* mutants mentioned previously. These mutants effected morphogenesis and isotope labelling demonstrated changes in proline catabolism in the mutant strain failing to use both arginine and proline as carbon and nitrogen sources (Han et al., 2019a). Fragment spectrums of metabolites can be utilised to distinguish between analytes with same *m/z* value for improved resolution of untargeted metabolomics. Fragmentation patterns of peaks can be compared to known spectrums to determine which analyte is present with the sample. This has also been used to infer greater detail from untargeted metabolomics in biofilm and in particular *Candida* biofilm research (Burgain et al., 2020).

1.5.7 Proteome

In a similar fashion to metabolomics, shotgun proteomics has progressed to become a popular and effective method of identifying and characterising microbial populations. Proteomics like other omic technologies aims to provide a holistic and complete data set of proteins within a biological system at a given point in time. LC-MS has been employed for example within *C. albicans* biofilms to identify many proteins in a high throughput manner (Shibasaki et al., 2018). In LC-MS the proteins undergo enzymatic cleavage, such as that performed by trypsin, into smaller peptides. These peptides are then ionised and mass spectrometry is performed by one of the MS platforms such as the Q-Exactive Plus (Thermo Fisher Scientific) before data acquisition of peptides by software such as the MASCOT searching engine (Karpievitch et al., 2010). These digested peptides are then used to infer the presence and quantity of proteins present in the biological sample.

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

Proteomics by LC-MS has been effective at characterising *Candida* dynamics in response to several conditions. Specifically, proteomics has been employed to decipher the physiological changes in *C. albicans* in the presence of serum over time (Shibasaki et al., 2018). Based upon an earlier study this analysis was performed within a type of strain to further elucidate the serum enhanced hyphal morphogenesis. They discerned that up to 50 serum induced protein changes were observed related to detoxification of oxidative species, glucose transport, TCA cycle and iron uptake (Aoki et al., 2013). Within the study mentioned here they considered proteomics an effective tool for identifying antigens for vaccines and therapeutic targets for drugs. In addition, proteomics by LC-MS has been performed to identify comparative differences between *Candida* species such as *C. glabrata* and *albicans*, host evasion mechanisms by *C. albicans* and *C. albicans* response to stress (Taff et al., 2012, Arita et al., 2019, Jacobsen et al., 2018). Shotgun proteomics has advanced *Candida* proteome research, as it has done for many biological systems, where previously we used lower throughput systems such 2D-PAGE (Shibasaki et al., 2018).

The fungal cell wall plays a pivotal role in biofilm development, hyphal elongation, adhesion, and host invasion. Due to this the composition and the changing dynamics of the cell wall are of particular interest. As discussed earlier, many cell wall bound adhesions and enzymes are present that both aid in biofilm formation and virulence (de Groot et al., 2004). For fungal cell wall proteomics homogenisation of the cell wall must first be performed, followed by centrifugation and multiple washes in NaCl. Then crude cell walls are then extracted with SDS-mercaptoethanol buffer. Digestion and mass spectrometry can then be performed in the same fashion as mentioned early without cytoplasmic and none cell wall bound contaminants (Dutton et al., 2014). In a 2014 study involving *O*-mannosylation deficient *Candida* mutants, a process of adding mannose to glycan chains, were investigated using cell wall proteomics. This process was previously determined to be important in hyphal formation and fungal interaction with bacterial species. Through cell wall proteomics they were able to determine perturbations in cell wall protein composition due knocking out the mannosyltransferase genes *Mnt1* and *Mnt2*. Many adhesion and virulence related cell wall proteins such as the ALS adhesin family of proteins and the Sap candidapepsin-9

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

precursor were reduced due to inhibition of mannose transfer (Dutton et al., 2014). In a more recent study performed in Aberdeen, the influence of the antifungal agent caspofungin was performed. They investigated the specific fungal cell wall proteome changes that were influenced across several *Candida* spp. including *C. albicans* in response to the echinocandin caspofungin, identifying changes such as the conserved expression of the glycosidase Utr2 involved in crosslinking chitin and surface glucans. Echinocandins inhibit the correct synthesis of β -glucan in the fungal cell wall. Overexpression of the Urt2 was assumed to be a compensatory mechanism due to the inhibitory effect of caspofungin on glucan production (Walker and Munro, 2020).

1.5.8 Integrative Analysis

The systems biology approach to microbiology involves integration of the sum of microbiome, transcriptome, proteome, and metabolome data sets. This potentially allows for a more comprehensive view of the system of interest than just one individual technology would allow for. Integration of multiple omics data sets has been demonstrated to be a useful tool (Zhang et al., 2010). With the increase in data analysis tools and increase in computational power, methods for multiple level integration of datasets are now possible.

In metabolomics and transcriptomic integration for example correlation analysis, following both data sets being processed allows for the identification of relevant metabolites and genes (Cavill et al., 2016). Network and pathway analysis have also become popular methods for integrating and interpreting different omics data sets (Liu et al., 2017, Fukushima et al., 2014). Numerous software packages have been created from online interfaces (Cottret et al., 2010) to packages developed for the R programming language (Luo et al., 2017). These packages allow for identification of overlapping enriched pathways in the transcriptome and metabolome data.

Conceptual integration through separate analysis of two or more omics datasets are perhaps the simplest to implement and the first stage in data integration. Through our subjective observations of the outcomes of separate analysis of two data sets it is possible to identify changes that are consistent or complimentary across two parallel

datasets (Cavill et al., 2015, Fox et al., 2014). Pathway analysis is becoming a more robust method for interrogating omics data due to the increase in *a priori* knowledge that is continually being curated for an increasing number of species. This has led to the development of network and pathway integration tools such as MetExplore and PaintOmics which are able to accept data from multiple omics datasets (Cottret et al., 2010, Hernández-de-Diego et al., 2018). They can overlay and map linked pathways between two datasets to identify features represented by both transcripts and metabolites. Enrichment methodologies additionally allow for significance to be determined based upon the representation of features from different omics (Gloaguen et al., 2017, Chong et al., 2018a).

Multivariate and dimensionality reduction techniques have also been shown to make significant advancements in data integration. Both supervised and unsupervised methods have been developed. Unsupervised methods cluster the data into different groups based upon variability within the data without any external guidance (Bouhaddani et al., 2018). Supervised analysis in contrast through regression analysis such as in the case of partial least squares (PLS) based techniques allow for the identification of important and key features or variables which distinguish between categorized data (Rohart et al., 2017, González et al., 2012). Examples include identification of important features in metabolomics and transcriptomic data sets that are related to chemical exposure in liver cells (Mesnage et al., 2018). Discriminatory covariate features in the microbiome and metabolome are also possible to infer through a discriminatory version of PLS discriminatory analysis (PLS-DA). For example in faecal samples, by integrating metabolomics and the microbiome data the authors were able to identify bacteria that were correlated with metabolic alterations in the gut of Crohns' disease sufferers (Metwaly et al., 2020).

1.6 Concluding Remarks and Aims

Biofilms as the mediators of disease are commonplace, with interkingdom interactions posing a difficult relationship to be fully understood. There are fluctuations in pathogenicity which are due to the varying relationships of different organisms,

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

environmental factors, pH of the host and genetic predisposition. *Candida* species live in complex biofilm communities, in particular *C. albicans* is of special note, being present at many clinically relevant sites. The species-species, kingdom-kingdom and even microbe-host relationships are extremely complex. It is due to this complexity that there is a need for a move towards holistic high throughput methodologies, RNA-Seq and metabolomics. These approaches offer an unbiased, large scale view which can lead to the discovery of medically important biomarkers, functional pathway identification in disease and potential intervention points. An understanding of the factors which change a microbial system from an inert system to that of a disease forming one is a necessity and will ultimately allow for prophylactic methods and therapeutic targets to be discerned.

Throughout the work within this thesis, we aim to utilise omics and combinations of omics platforms to interrogate heterogenous *C. albicans* biofilms. Previously, it has been discussed the complexities, the clinical relevance and the advancements that have motivated and enabled the work within this thesis. We hypothesise that we can improve our understanding of biofilm formation through these microbiome, transcriptome and metabolomic techniques. We also wish to validate these methods and similarly identify future biofilm models on which they can be utilised. We aim to comprehensively identify future complex-biofilm models and methods of holistic omics methodologies to which to apply to these models. Specifically, the work contained within this thesis aims to:

1. Identify interactions with *Candida* in microbiome data from a recent denture stomatitis clinical study. We endeavour to reanalyse this clinically relevant denture microbiome study with a close focus on the impact of *Candida albicans* on the oral microbiome. We also aim to identify the impact of oral hygiene on both *Candida* and the bacterial microbiome through interrogation of microbiome data.

Chapter 1: *Candida* biofilms, their clinical significance and role in interkingdom interactions

2. Utilise omic technologies to identify:
 - A. Differentially expressed genes using RNA-Seq to profile the transcriptome of phenotypically distinct isolates. Additionally, how their profile changes in response to biofilm inducing stimulus.
 - B. The changes in the composition of the cell wall proteome between phenotypically distinct isolates. Additionally, how the composition of the cell wall proteome is impacted in response to biofilm inducing stimulus
 - C. The metabolic footprint from phenotypically distinct isolates in the presence and absence of biofilm inducing stimulus to determine changes in metabolic function.

The commonly used foetal calf serum will serve as our biofilm inducing media throughout these experiments.

3. Identify methods and analysis techniques that are available in the literature which can integrate omics data to assess the applicability of these methodologies at integrating metabolomic and transcriptomic datasets. Finally, we aim to determine the usefulness and interpretability of integrated data from microbiological datasets.

2 The impact of *Candida* and its interactions with the microbiota at the denture surface

2.1 Background

Denture stomatitis (DS) and denture related illness have historically been associated with and *C. albicans* and other members of the *Candida* genus. Indeed, it is common within microbiology for individual species to become heavily implicated in specific diseases. This is true of *Candida* spp. in the case of DS where clinical studies have demonstrated correlations with the quantities of *Candida* and the severity of the disease through the Newton's classification of disease severity (Coco et al., 2008b, Bastiaan, 1976). It has been further reported that the individual strain of *C. albicans* and its relative capacity to form biofilms plays a role in disease outcome in denture-related disease (O'Donnell et al., 2017); a characteristic trait that has also been shown to be true in other systemic diseases (Kean et al., 2018a, Rajendran et al., 2016d, Sherry et al., 2017). Moreover, it has been shown that strain specificity also has a bearing on denture cleansing capacity, with those individuals harbouring more prolific biofilm forming strains negatively responding to treatment (O'Donnell et al., 2017). Taken together, these studies support the dogma of a simple mono-species oral infection. This has led to the disproportionate focus on *Candida* spp., particularly related to assessing denture oral hygiene strategies *in vitro* and *in vivo* (Nunes et al., 2016, Ramage et al., 2012c).

Dentures are generally bathed in the salivary microbiome, a microbial soup interacting on the surface of the denture acrylic. Denture plaque biofilms are complex and are often represented by dense, mixed interkingdom communities (Delaney et al., 2018). The quantities of yeast and bacteria found residing upon the pores and varied denture topography are vast yet are relatively poorly explored (O'Donnell et al., 2015b, Shi et al., 2016). The first molecular-based microbial denture studies revealed a complex bacterial microbiota with potentially cariogenic, periodontopathic and malodourous capacities (Mantzourani et al., 2010, Sachdeo et al., 2008, Yitzhaki et al., 2017). Moreover, in mixed dentition where there is the presence of natural teeth within a partially edentulous patient, differential ecology is observed that may also have a bearing in the progression of oral disease (Teles et al., 2012). Despite these studies providing greater insights into the complexity of oral bacterial biofilm ecology, they

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

fail to account for the involvement of *Candida* spp. within the community. We know these yeasts are prone to contribute to a less diverse biofilm (Cho et al., 2014, O'Donnell et al., 2015b), which can also play a leading role in driving denture-related stomatitis.

With this in mind and based on our greater understanding of the importance of bacteria and yeasts within the denture biofilm, investigating both the bacterial and yeast role during the development and testing of oral hygiene products is important. The Ramage laboratory were the first in 2017 to undertake a randomised double blinded control trial to assess the importance of frequency of denture cleaning (Ramage et al., 2019), where both bacteria and yeasts were quantified as primary outcome measures. This study statistically demonstrated the benefit of frequent (daily) cleansing regimens compared to intermittent regimens. Nevertheless, a key limitation of the study was the failure to employ next generation sequencing techniques to fully assess the microbial composition of patients under different cleaning regimens. Given the importance of both bacteria and yeasts within the denture environment, in addition to the impact of oral hygiene on DS, this study is the first seeking to understand how the oral microbiome is impacted by self-reported oral hygiene behaviours.

2.2 Hypothesis and aims

This component of the thesis hypothesised that denture cleansing hygiene practices influenced the presence of yeasts. The following chapter therefore examines the contribution of oral hygiene measures and the relevance of *Candida* spp. to the denture microbiomes of edentulous patients, with the overarching aim of improving denture antimicrobial strategies.

Specifically, the study aims to identify:

- Perturbations in the oral microbiome, within the dental plaque, mucosa, and denture plaque, which are due to differences in the oral hygiene of the wearer

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

- Perturbations in the oral microbiome within the dental plaque, mucosa, and denture plaque, which are due to the different sleeping habits within our patient cohort.
- Changes in the oral microbial communities that are influenced by the presence or prevalence of *Candida* spp. within the dental plaque, mucosal and denture plaque communities.
- Specific genera that are related or in higher abundance between our different patient demographics.

Overall, any variances in the overall community, and any specific changes, that are influenced by the hygiene of the denture wearer or the proportion of yeasts, will be explored and interrogated in detail within this chapter

This chapter has been published in the Elsevier journal Biofilm under the title “Interkingdom interactions on the denture surface: Implications for oral hygiene”- <https://doi.org/10.1016/j.bioflm.2019.100002>. Additionally, data from this chapter has been presented at the annual Oral Microbiology and Immunology Group (OMIG) conference.

2.3 Methods

2.3.1 *In vitro* denture cleansing study

A denture plaque cleansing study and quantitative analysis of remaining viable cells was performed according to an established methodology devised within the laboratory (Sherry et al., 2016a). It was the sub-aim to investigate the impact of a frequent denture cleansing regimens on bacterial and yeast retention on denture acrylic, with an aim to understand how complex interkingdom biofilms respond to oral hygiene measures. Specifically, this was a combined chemical and mechanical brushing cleansing technique, employed sequentially over a 7-day treatment period.

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

Briefly, laboratory strains were used to create a polymicrobial denture plaque biofilm model based on the most dominant genera/species identified from our recent denture microbiome study (O'Donnell et al., 2015b, Ramage et al., 2019). Polymethylmethacrylate (PMMA) discs were manufactured by Dr Hasanain Alalwan (PhD student), as previously described (Alalwan et al., 2018), which provided the physical substrates on which biofilms were formed. The biofilms included the following composition of bacteria and yeasts: *Streptococcus mitis* NCTC 12261, *Streptococcus intermedius* ATCC 27335, *Streptococcus oralis* ATCC 35037, *C. albicans* 3153A, *Actinomyces naeslundii* ATCC 19039, *Veillonella dispar* ATCC 27335, *Rothia dentocariosa* DSMZ 43762, *Lactobacillus casei* DSMZ 20011 and *Lactobacillus zeae* DSMZ 20178. Maintenance conditions on agar and overnight culture conditions for each of the organisms are outlined in Table 2.1.

Table 2.1 Maintenance and culture conditions for oral denture species

Oral Species	Agar	Broth	Growth Condition	Temperature
<i>Streptococcus mitis</i> NCTC 12261	Columbia Blood Agar (CBA)	Tryptic Soy Broth (TSB)	CO ₂	37°C
<i>Streptococcus intermedius</i> ATCC 27335	Columbia Blood Agar (CBA)	Tryptic Soy Broth (TSB)	CO ₂	37°C
<i>Streptococcus oralis</i> ATCC 35037	Columbia Blood Agar (CBA)	Tryptic Soy Broth (TSB)	CO ₂	37°C
<i>C. albicans</i> 3153A	Sabouraud's Dextrose Agar	Yeast Peptone Dextrose (YPD)	O ₂	30°C
<i>Actinomyces naeslundii</i> ATCC 19039	Fastidious Anaerobic Agar (FAA)	Brain Heart Infusion (BHI)	AnO ₂	37°C
<i>Veillonella dispar</i> ATCC 27335	Fastidious Anaerobic Agar (FAA)	Brain Heart Infusion (BHI)	AnO ₂	37°C
<i>Rothia dentocariosa</i> DSMZ 43762	Columbia Blood Agar (CBA)	Brain Heart Infusion (BHI)	O ₂	37°C
<i>Lactobacillus casei</i> DSMZ 20011	MRS Agar	MRS Broth	CO ₂	37°C
<i>Lactobacillus casei</i> DSMZ 20011	MRS Agar	MRS Broth	CO ₂	37°C
<i>Lactobacillus zeae</i> DSMZ 20178	MRS Agar	MRS Broth	CO ₂	37°C

Initially, *S. mitis*, *S. intermedius*, *S. oralis* and *C. albicans* were grown and standardised in artificial saliva (AS) to 1×10^7 cells/mL. These were added to each well of a 24 well plate (Corning Inc, New York, USA) containing 13 mm² PMMA discs (Chaperlin and Jacobs Ltd, Southend-On-Sea, UK) and incubated aerobically at 37°C for 24 h. Next, standardised (1×10^7 cells/mL) *A. naeslundii*, *V. dispar*, *R. dentocariosa*, *L. casei* and *L. zeae* were added to the preformed 24-h biofilm and incubated at 37°C in 5% CO₂ conditions for a further 4 days. Spent supernatants were removed and replaced with fresh AS daily. AS was comprised of porcine stomach mucins (0.25% w/v), sodium chloride (0.35% w/v), potassium chloride (0.02 w/v), calcium chloride dihydrate (0.02% w/v), yeast extract (0.2% w/v), lab lemco powder (0.1% w/v), proteose

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

peptone (0.5% w/v) in ddH₂O (Sigma-Aldrich). Urea was then added independently to a final concentration of 0.05% (v/v).

The treatment regimen was a daily treatment (days 1-7) of a 3min soaking with a denture cleanser (Polident® 3-minute denture cleanser; GSK Consumer Healthcare, Weybridge, UK) followed by brushing with hard water. For analyses, sample biofilms were assessed pre- and post-treatment. Following each treatment, PMMA discs were incubated in Dey-Engley neutralising broth (Sigma-Aldrich, Gillingham, UK) for 15min. PMMA discs were then sonicated in phosphate buffered saline (PBS [Sigma-Aldrich, Gillingham, UK]) at 35 kHz for 10min to remove the biomass, as previously described (Ramage et al., 2012c). The treatment regimen is visualised in Figure 2.1.

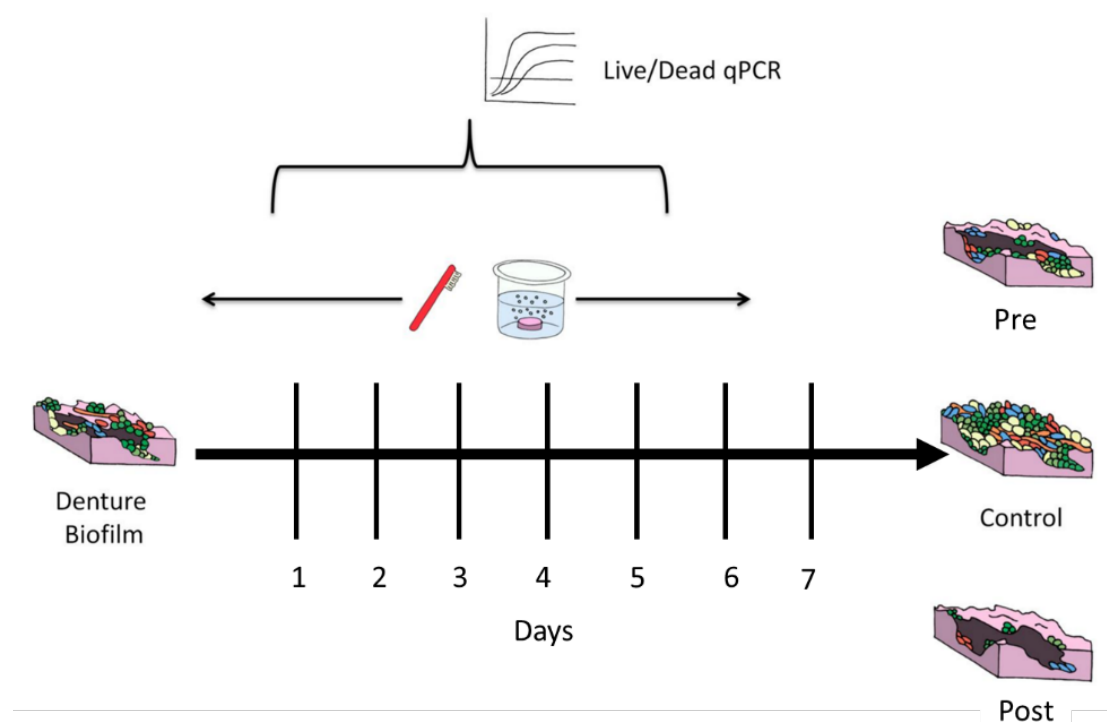


Figure 2.1 Treatment regimen for invitro assessment of denture cleansing on bacterial and fungal loads. The denture plaque biofilm model was prepared on PMMA discs. PMMA discs were prepared so that each day a pre and post treatment as well as an untreated PMMA disc could be taken. PMMA discs for subsequent days were also chemically and mechanically treated. PMMA discs control, pre and post treatment were prepared every day for the 7 days and sonicated before quantification of colony forming equivalents before qPCR.

For quantitative analysis, quantitative live/dead PCR were performed, as described previously (Sherry et al., 2016a), and detailed below. Live/dead PCR was performed

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

using 16S and 18S bacterial and fungal specific primers, and quantified using appropriate bacterial and yeast standard curves (O'Donnell et al., 2015c).

2.3.2 *In vivo* denture hygiene study

The clinical study specimen collection was initially performed by Dr Lindsay O'Donnell. I can confirm that my role in this study was to undertake detailed microbiome analysis pipelines to interrogate and integrate the relevant meta-data. Study collection through to analysis is illustrated in Figure 2.2.

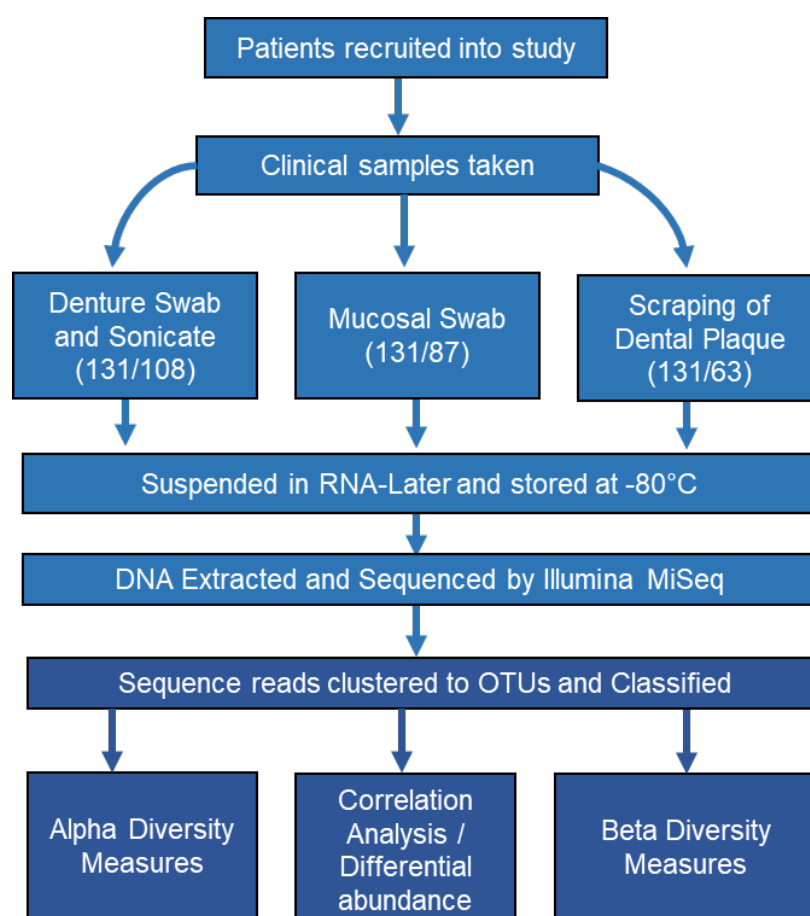


Figure 2.2 **Sample collection and Analysis Workflow.** From 131 denture wearers a denture swab, mucosal swab and dental plaque scraping were taken. Swabs were sonicated and all samples were stored in RNA-Later. All samples from 131 patients underwent DNA extraction for sequencing due to failures in extraction and library prep 108 denture samples, 87 mucosal samples and 63 dental samples remained for sequencing. After sequencing samples reads were processed for analysis. Microbiome data was obtained from previously published study and full experimental details are published there. The original work was carried out by Lindsay O'Donnell at the Glasgow Dental Hospital and School (O'Donnell et al., 2015b).

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

2.3.3 Study participants

131 denture wearing patients attending the University of Glasgow Dental Hospital and School were enrolled in the study, as described previously (O'Donnell et al., 2015b). Written informed consent was obtained from all participants. Ethical approval for the study was granted by the West of Scotland Research Ethics Service (12/WS/0121). Clinical assessments were carried out by six experienced dentists working in the prosthodontic department of the University of Glasgow Dental Hospital and School. All prosthodontists received personal training from Dr Douglas Robertson (Senior Clinical Lecturer in Restorative Dentistry and study Principal Investigator) to standardise the assessment of the clinical disease (inflammation), denture retention, stability, occlusion and cleanliness.

Oral hygiene was graded after training and discussion as *good* (with little or no signs of denture or dental plaque visibly present), and *poor* (generalised or gross denture or denture plaque evident). All examiners were trained but no formal calibration calculations were carried out. The patient demographic and clinical examination data was recorded on a standardised data collection sheet. There were no age-related exclusion criteria for this study.

There were several exclusion criteria included in this study which included the exclusion of pregnant women, those that had undergone radiotherapy for head and neck related malignancies, those with periodontitis and those using any prescribed antimicrobial therapies, prescription mouthwashes or immunosuppressants within six months of the study.

2.3.4 Clinical sample collection

Ethylene oxide sterilised swabs (Fisher Scientific, Loughborough, UK) were used to take samples from the denture surfaces in contact with the palatal mucosa and the palatal mucosal surface covered by the dentures. Samples were collected and processed as previously described (O'Donnell et al., 2015b). In total, samples from 131 patients were collected, which included 131 denture swabs and 131 mucosal swabs. However, during DNA extraction process not all samples had sufficient DNA, and

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

therefore only DNA from 108 denture samples, 87 mucosal samples and 63 dental samples remained for sequencing, collectively all these samples originated from 123 patients.

In parallel, dentures removed from the patients' mouths were placed in sterile bags (Fisher Scientific) containing 50 ml PBS (Sigma-Aldrich, Dorset, UK). Adherent denture plaque was then removed by sonication (Ultrawave, Cardiff, UK) for 5min, as previously described (O'Donnell et al., 2015b). Bacterial and fungal loads were quantified by qPCR using 16S and ITS primers, as described previously (Kraneveld et al., 2012, O'Donnell et al., 2015b).

2.3.5 DNA Isolation

All samples were prepared for DNA isolation as previously described (O'Donnell et al., 2015b), using a combination of chemical and mechanical lysis. DNA was isolated from clinical samples by first either centrifugation at 13000g for 15min for dental plaque samples before resuspension in 150ul of TE buffer, or sonication for swab samples before the same treatment. All samples were then lysed within a plate containing 0.24ml of lysis buffer 0.25 ml of lysis buffer (AGOWA mag Mini DNA Isolation Kit, AGOWA, Berlin, Germany), 0.3g zirconium beads (diameter, 0.1mm; Biospec Products, Bartlesville, OK, USA) and 0.2 ml phenol. The samples within the plate were then homogenised using the Mini-beadbeater (Biospec Products) for 2min (O'Donnell et al., 2015b). Bacterial and fungal loads were quantified by qPCR using 16S and ITS primers, as described previously (Kraneveld et al., 2012, O'Donnell et al., 2015b). Primers and a probe for the 16S rRNA gene were used and primers for the *Candida* ITS genes utilised to determine *Candida* colony forming equivalents (CFEs) are shown in Table 2.2.

Table 2.2 Primers for qPCR and Illumina Sequencing.

Target	Primer Sequence	Reference	Method
16S (Bacteria)	F - TCCTACGGGAGGCAGCAGT R - GGACTACCAAGGTATCTAATCCTGTT Probe - 6FAM-CGTATTACCGCGGCTGCTGGCAC-BBQ	Kraneveld et al., 2012	qPCR
ITS (Fungi)	F - CCTGTTTGAGCGTCRTTT R - TTCTCCGCTTATTGATAT	Kraneveld et al., 2012	qPCR
V4 region	F 515F -GTGCCAGCMGCCGCGGTAA R 806R - GGACTACHVGGGTWTCTAAT	Caporaso et al., 2011	Illumina

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

A total reaction volume of 20 μ l was used containing 3 μ l of DNA. Reactions contained 2x PCR Probe Master Mix (Roche) for 16S or 2 X SYBR Green PCR Master Mix (Roche) in the case of ITS primers. The conditions of the qPCR were activation for 10min at 95°C followed by 50 cycles of denaturation for 30secs at 95°, annealing for 40secs at 60°C and extension for 30secs at 75°C. qPCR was carried out in the LC480-II light cycler (Roche) before bacterial 16S rDNA concentrations (CFE) were determined from standard curves of *E. coli* K12 cultures and fungal ITS concentrations were extrapolated using *Candida dubliniensis*.

2.3.6 Illumina Sequencing

For each individual sample amplicon libraries of the V4 hypervariable region of the 16S rRNA gene were generated. The amplification mixes and PCR conditions used were as previously described (O'Donnell et al., 2015b). Briefly, amplicons from the V4 hypervariable region were prepared, and these primers also included Illumina adapters and unique 8nt long barcode sequences. The libraries were amplified using 2 units of Phusion HotStart II High fidelity polymerase (Thermo Scientific), 1-unit Buffer Phusion HS II [5x], including 1.5 mM MgCl₂ (Thermo Scientific), 0.2 mM dNTP (Thermo Scientific, Germany) and 1 μ M of each primer. After denaturation (98°C; 30sec), 35 cycles of denaturation (98°C; 10sec), annealing (55°C; 30sec), and extension (72°C; 30sec) were performed. DNA libraries were pooled in equal amounts before the being purified using the Illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, Eindhoven, the Netherlands). The amplicon was sequenced in paired end mode on a MiSeq sequencing system (Illumina, Eindhoven, the Netherlands) with the v2 kit (Illumina) (Caporaso et al., 2012, Kozich et al., 2013).

2.3.7 Sequencing data analysis

Reads were first quality filtered using Trimmomatic v0.32, (Bolger et al., 2014). Next, the reads were merged using fastq-join implemented in QIIME v.1.8.0 (Bolger et al., 2014). Sequences were clustered into operational taxonomic units (OTUs) using USEARCH v7.01090 (Edgar, 2013), after quality filtered with usearch (maxee 0.5). The representative sequence of each cluster was assigned a taxonomy using the RDP classifier (Cole et al., 2009). The Ribosomal Database Project: improved alignments

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

and new tools for rRNA analysis. Nucl Acids Res 37: D141–145. doi: 10.1093/nar/gkn879) (QIIME v.1.8.0) (Greengenes v13.8 97_otus set) with a minimum confidence of 0.8.

2.3.8 Study design and statistical analysis

The study was designed as a pilot study and was initially only powered to detect a biologically meaningful association between diseased or healthy mouths and microbiome composition, and therefore was not originally powered to detect differences between additional variables including, denture hygiene, cleaning frequency and sleeping habits.

The data set was randomly sub-sampled to 770 reads per sample (minimum number of reads per sample was 776), the use of a rarefaction curve based on our data deemed that 770 reads per sample was sufficient to avoid minimal loss in diversity which allowed for inclusion of the maximum number of samples for analysis. Statistical analysis was performed within R on the OTU, and taxonomic tables created, as described above. Additionally, meta table data from clinical parameters and other *in vitro* analysis including CFEs from qPCR were used for analysis within this chapter. Community analysis was performed using the both the *Simpson* and *Shannon* alpha diversity indexes within the phyloseq package (McMurdie and Holmes, 2013). Diversity indexes aim to apply a numerical quantification to the level of species sample within a community. *Simpson's* index gives a measure of dominance in which the value increases the level of dominance increases and therefore diversity is said to decrease. The *Shannon* index increases as both the richness and evenness of the community increases.

Nonmetric Distance Scaling (NMDS) plots of community data were performed using *Bray-Curtis* distances on community data represented as OTUs. Additionally, the distance measure based upon *Weighted UniFrac* was implemented using the phyloseq package. UniFrac is a β -diversity measure which unlike other distance metrics such as bray Curtis incorporates phylogenetic information to discern whether communities differ in composition. Weighted version of UniFrac incorporates the abundance of

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

each feature so branches of the tree are weighted by the proportional abundance of each feature or taxa. This weighting removes the issue of low abundance taxa with long phylogenetic branches having unproportionate effect on the total distance calculated within the community. Non-metric multidimensional scaling (NMDS) ordination plots were constructed using one of the above distance metrics.

To measure OTUs that significantly differed between conditions, we used the R package DESeq2 (Love et al., 2014), with an adjust p -value of <0.05 and log2 fold-change cut off >1.5 . DESeq2 was originally developed for identifying differential expressed transcripts/genes in RNA-Seq transcriptome studies. The same principles can be applied to identify features, in this case taxa, that are in higher abundance in one condition compared to another. Additionally, correlation analysis and analysis of variance (ANOVA) were performed using function within base R. ANOVA was utilised to compare differences in diversity measures, *Shannon*, or *Simpson*, between our different communities. When measuring significance with ANOVA a p -values of <0.05 was used unless stated otherwise.

When analysing the experimental treatment data distribution, graph production and statistical analysis were performed using GraphPad Prism (version 5; La Jolla, CA, USA). After assessing whether data conformed to a normal distribution, One-way Analysis of Variance (ANOVA) and t tests were used to investigate significant differences between independent groups of data that approximated to a Gaussian distribution. A Bonferroni correction was applied to the p value to account for multiple comparisons of the data.

2.4 Results

2.4.1 Influence of denture hygiene on *in vitro* biofilms

In vitro denture biofilms were exposed to a combination of chemical and mechanical denture cleansing sequentially over a 7-day period. Live cell analysis was carried out by qPCR using the 16S and 18S rDNA primers. A control arm with no intervention was included for comparison. Figure 3.3 shows that denture cleansing was able to

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

significantly reduce the viable bacterial colony forming equivalents (CFEs) by at least 1 log₁₀ when comparing pre-treatment to post-treatment samples on each day of analysis ($p < 0.01$). However, when the *Candida* were quantified, we could not detect any significant difference in cell numbers within the treatment group pre- and post-denture cleansing, despite highly significant differences between the treatment arm and control group from days 3–7 ($p < 0.001$). Figure 2.3 shows that total bacteria and *Candida* CFEs are higher in all control and treatment arms, suggesting dead cells make up a considerable component of the dead biofilm. Taken together, these data indicate that bacteria within the biofilm are more sensitive to denture cleansing than *Candida*, alternatively, bacterial numbers during regrowth can suppress the ability of retained *Candida* cells to repopulate. Irrespective, significant numbers of *Candida* and bacteria are retained despite intense and frequent treatment regimens, suggesting some co-operative protection or tolerance within an interkingdom biofilm.

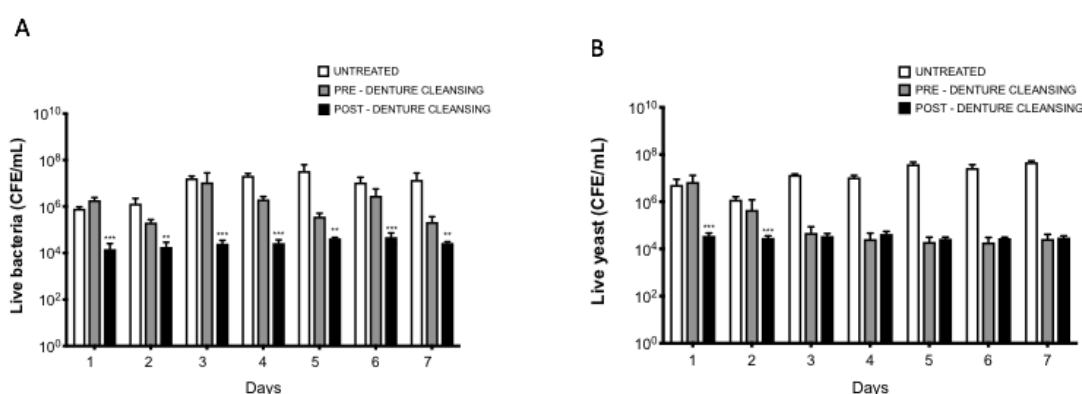


Figure 2.3 Quantitative live assessment of bacterial and *Candida* load following *in vitro* denture hygiene. Nine species of denture biofilms were grown on PMMA sections, followed by daily denture cleanser and brushing. Sonicate samples were taken pre- and post-treatment, and an untreated positive control included. Live cell numbers were determined by qPCR or PMA treated samples (live cells) of (A) bacteria, and (B) *Candida*. Data was analysed by ANOVA with a Bonferroni correction (** $p < 0.01$, *** $p < 0.001$).

One hundred and thirty-one (131) patients were recruited to this study, of which the primary demographics of these patients are shown in Table 2.3. The average patient age was 70.2 ± 11.5 years (min: 33, max: 95) with an average denture age of 4.5 ± 5.1 years (min: 0.2 max: 40). In terms of gender, females represented the majority of the population at 64.9%, with males contributing only 35.1%. After clinical diagnoses,

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

62.6% of participants were found to have healthy oral mucosa and the remaining 37.4% were diagnosed with DS. *These data were collected by Dr Robertson and Dr O'Donnell.*

Table 2.3 **Patient denture hygiene demographics.**

	Healthy	Disease
n	82 (62.6%)	49 (37.4%)
Male	26 (31.7%)	20 (40.8%)
Female	56 (68.3%)	29 (59.2%)
Mean Age	71.6	68.2
Median Age	72	70
Mean Age	4.5	4.4
Complete dentures	61 (74.4%)	28 (57.1%)
Partial dentures	21 (25.6%)	21 (42.9%)
Denture Hygiene		
Satisfactory	65 (79.3%)	30 (61.2%)
Poor	17 (20.7%)	19 (38.8%)
Denture cleaning		
once/day	24 (30.0%)	18 (37.5%)
≥ twice/day	56 (70.0%)	30 (62.5%)
Sleeping with denture		
No	43 (52.4%)	15 (30.6%)
Yes	39 (47.6%)	34 (69.4%)

Patients were classed as having either good or poor denture hygiene, in total 95 (72.5%) patients had good denture hygiene and 36 (27.5%) had poor denture hygiene. When separated into healthy and diseased groups, 20.7% and 38.8%, respectively, were classed as having poor denture hygiene. Denture cleaning varied among the cohort, however, the majority reported cleaning their dentures either once or twice per day. Forty-two (32.1%) participants reported cleaning their denture once per day, whereas 86 (65.6%) cleaned theirs twice per day. Going to sleep whilst wearing a denture is a habit that was commonplace amongst study participants, as 73 (55.7%) of the total patients reported that they sleep with their denture *in situ* (Table 2.2).

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

Additionally, patients were classified according to their oral health impact profile (OHIP) score calculated from the OHIP-14 questionnaire; each question had a 5-point scale to assess different aspects of oral health and quality of life. The sum of the questionnaire is used to assess their overall oral health with a maximum possible score of 56 (Allen and Locker, 2002).

2.4.2 Influence of denture hygiene on the denture associated *Candida*

The influence of *Candida* load on dentures was compared across several oral and denture hygiene practices, detailed in Figure 2.4. Regarding sleeping with the denture *in situ*, no statistical significance was observed between those who did and those who did not with respect to *Candida* load (Figure 2.4A). Whether the denture wearer had good, poor or excellent oral hygiene similarly appeared to have no effect on the *Candida* load (Figure 2.4B), nor did cleaning once or twice per day have an impact on *Candida* burden (Figure 2.4C). These data suggest that *Candida* species are not influenced by oral hygiene measures *in vivo*.

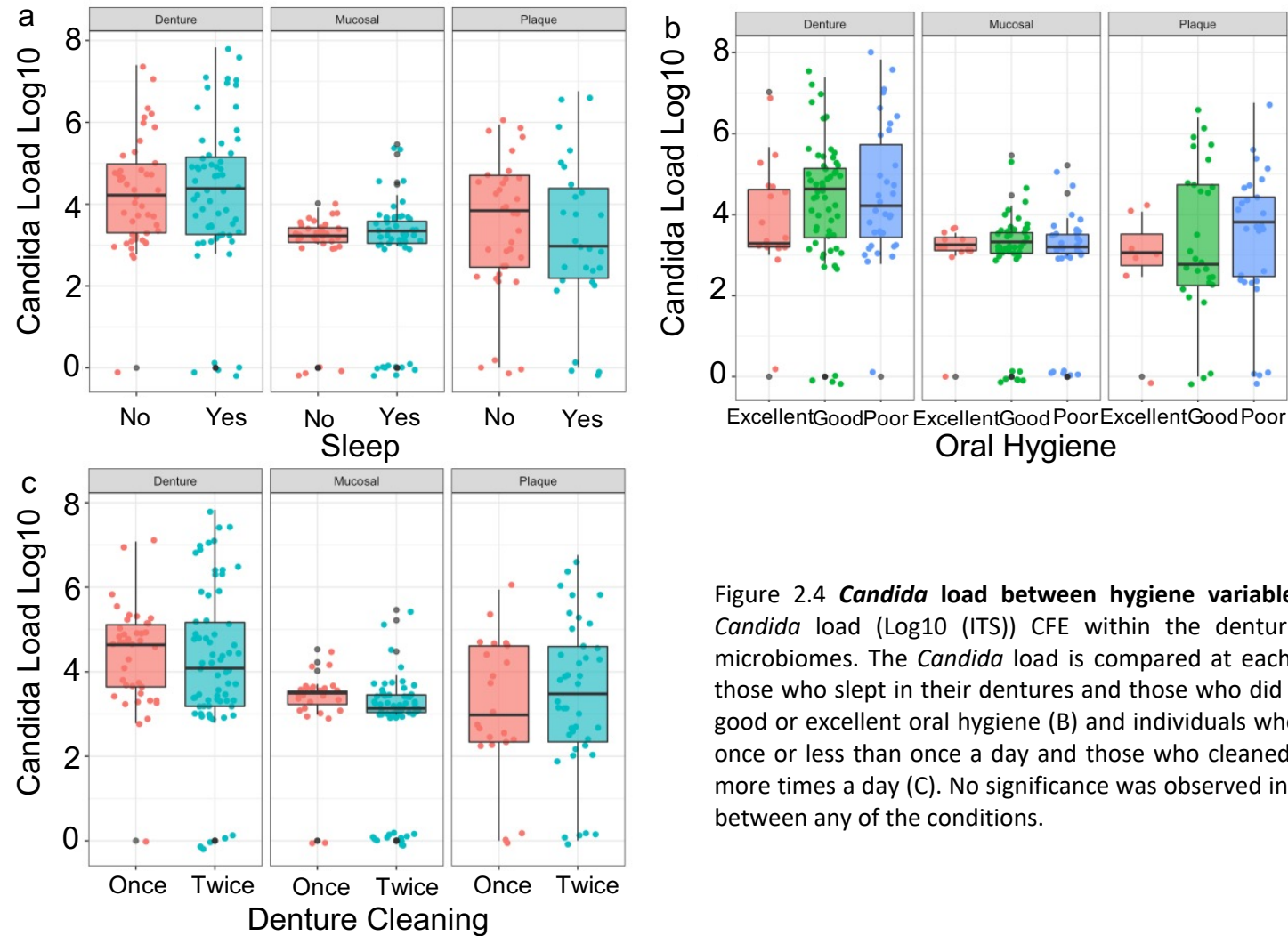


Figure 2.4 ***Candida* load between hygiene variables and oral sites.** Total *Candida* load (Log₁₀ (ITS)) CFE within the denture, mucosal and plaque microbiomes. The *Candida* load is compared at each of these sites between those who slept in their dentures and those who did not (A) those with poor, good or excellent oral hygiene (B) and individuals who cleaned their dentures once or less than once a day and those who cleaned their dentures twice or more times a day (C). No significance was observed in the overall candida load between any of the conditions.

2.4.3 Influence of denture hygiene on the oral microbiome

Ordination plots utilising the weighted UniFrac distance measure as well as ANOVA on the *Shannon* diversity and *Simpson* diversity index were implemented to compare the diversity and dominance of the denture microbiome for the different oral and denture hygiene practices. When using abundance and phylogenetic distances (Weighted UniFrac) to compare hygiene status, denture cleaning and sleeping with the denture *in situ* we observed no patterns in diversity between the different conditions (Figure 2.5). Similarly, when using the diversity metrics *Shannon* and *Simpson* to compare between conditions observed no significant differences. The diversity is not significantly affected by the hygiene status, denture cleaning frequency or whether an individual sleep in their denture within this cohort (Figures 2.6-2.8). These data suggest that bacterial species, composition, and diversity are not influenced by oral hygiene measures *in vivo*.

Despite there being no overarching consequence on the diversity and richness of the oral community between different hygiene measures, some individual changes in species abundance were observed with respect to oral hygiene status and those who did and did not leave their dentures *in situ* whilst sleeping (Figure 2.9). No significant changes in abundance of species were observed in the dental microbiome (Figure 2.9A), however the mucosal microbiome (Figure 2.9C) had 1 genus significantly represented (*Dialaster*) in those who slept with denture *in situ*. The denture microbiome (Figure 2.9E) had a number of genus (*Leptotrichia*, *Selenomonas*, *Moryella*, *Prevotella*) in significantly higher abundance in those who slept with their denture *in situ* compared to those who did not. Within the dental, mucosal and denture microbiome a number of species are in significantly higher proportion in those with poor oral hygiene (Figure 2.9). The bacterial genus *Scardovia* was in significantly higher abundance in all three microbiome sites (Figure 2.9B, D and F) in those who had poor oral hygiene. Those with poor oral hygiene also had higher levels of *Fusobacterium* and *Schwartzia* in the denture microbiome (Figure 2.9F), and increased levels of *Lactobacillus* within their dental microbiome (Figure 2.9B). These data indicate that specific genera of bacteria are influenced by oral hygiene practices.

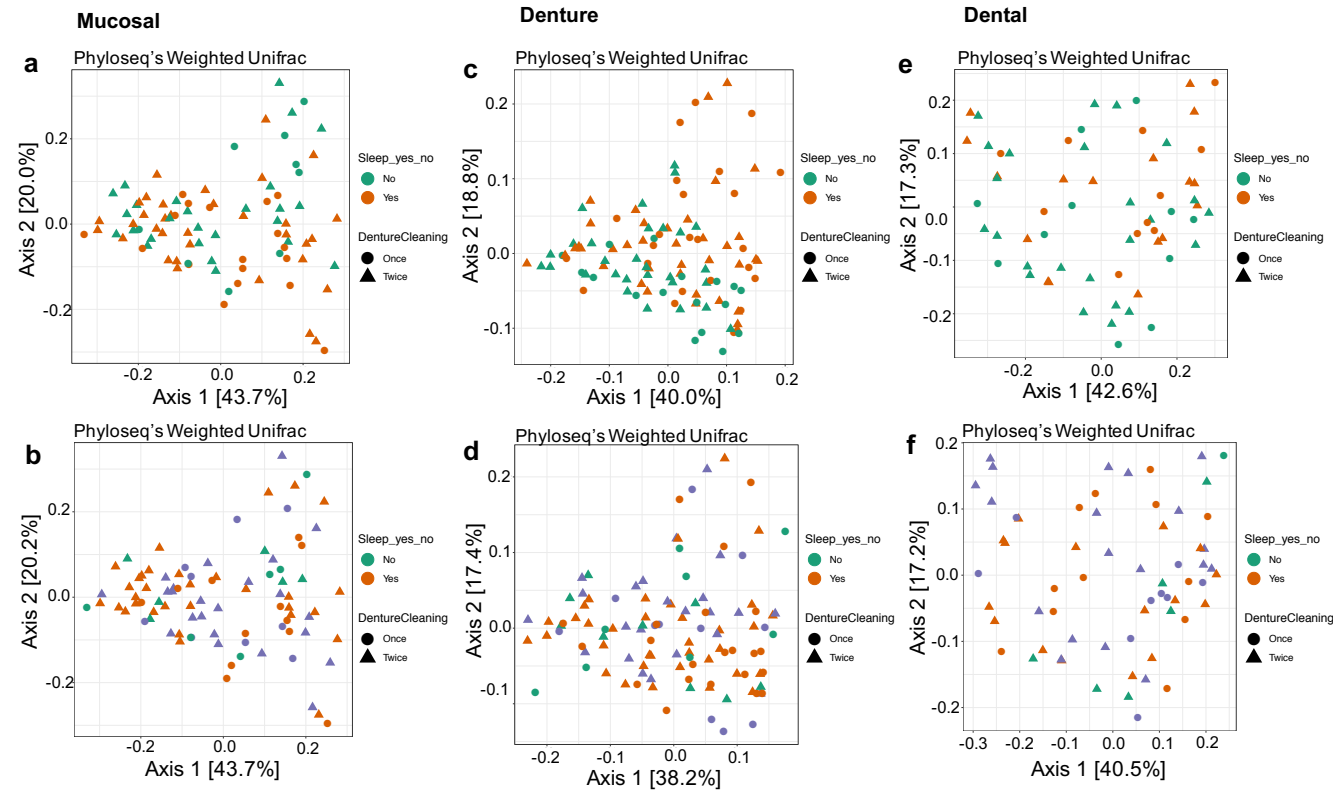


Figure 2.5 **PCoA ordination plot based on phyloseq's weighted UniFrac distances**. Principal coordinates analysis (PCoA) of bacterial diversity based on weighted UniFrac distances from the microbiome of the mucosal (a, b), denture (c, d) and the dental (e, f) surfaces. The patient variables: whether they sleep in their denture and their hygiene rating are coloured on individual plots. Additionally, the patient variable Denture cleaning, whether they clean their denture once a day/less once a day or twice a day/more than twice a day is represented by a shape. The largest amount of variation is visualised on the x axis with second largest on the y. Percentages represent the variance that is represented between samples on each axis.

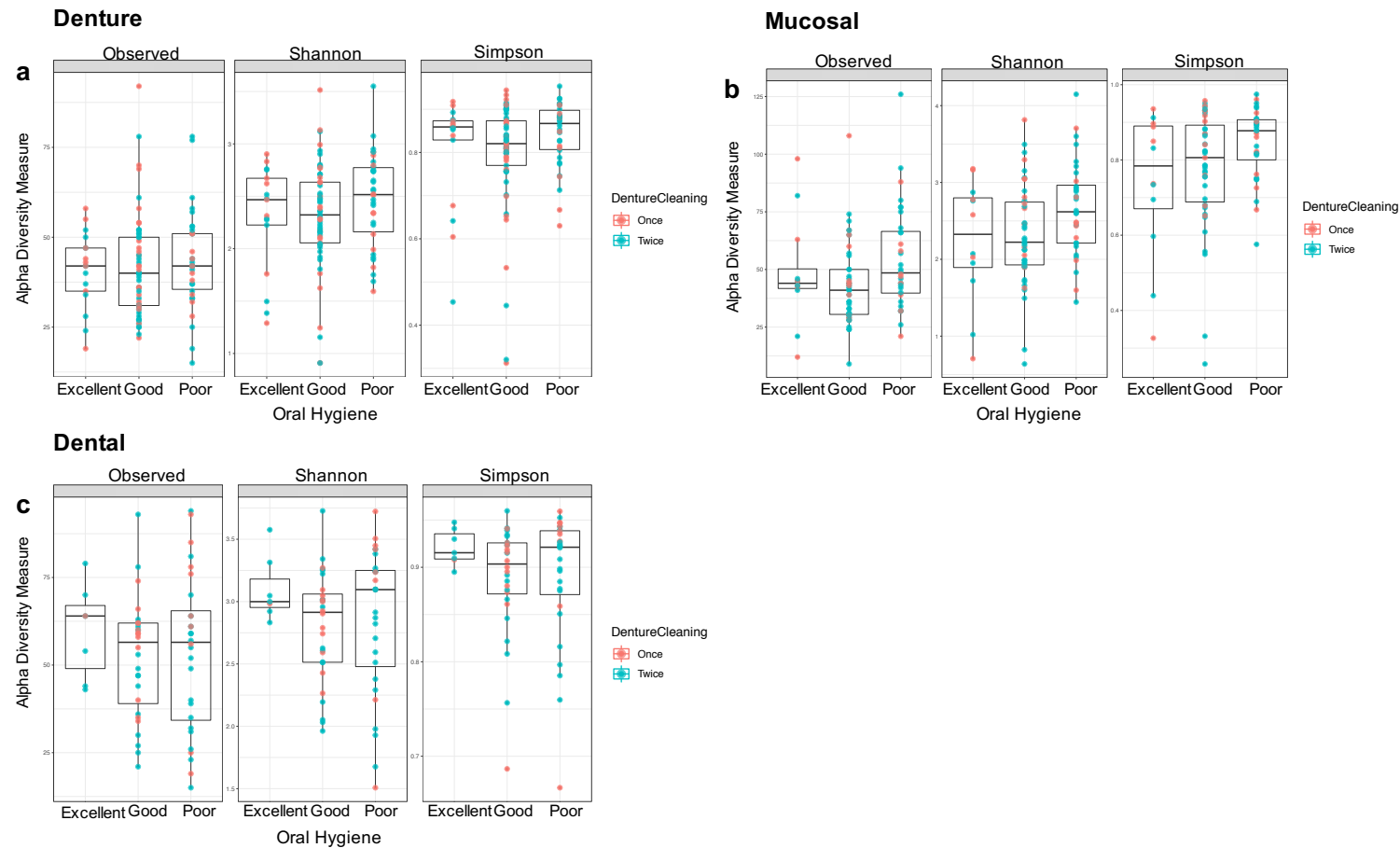


Figure 2.6 Microbiome diversity measures between poor, good and excellent oral hygiene. Diversity between patients with poor, good and excellent oral hygiene were measured for each of the three oral sites: denture (a), mucosal (b) and dental (c) using the observed OTUs, *Shannon* Index and the *Simpson* diversity index. The upper and lower quartiles are indicated by the top and bottom boundaries of the boxplots while the line within the plot indicates the median values. Analysis of variance was performed (ANOVA) between patients' variables for each site on each of the diversity measures and was only reported if $P < 0.05$. None of the variables fulfilled this criterion so no significance bars are shown.

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

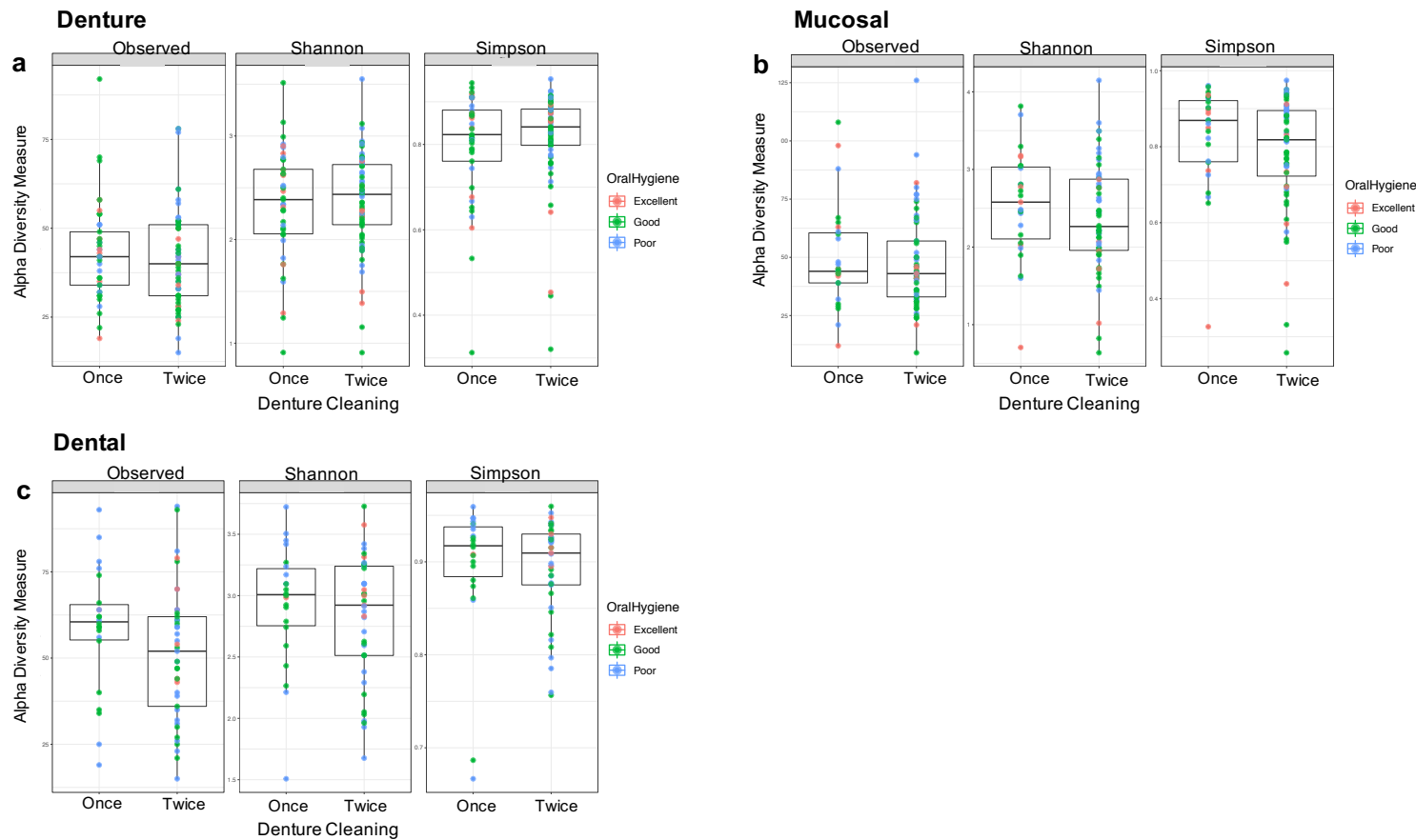


Figure 2.7 Microbiome diversity measures between denture cleaning once or twice. Diversity between patients who clean their dentures \leq once a day and \geq twice a day were measured for each of the three oral sites: denture (a), mucosal (b) and dental (c) using the observed OTUs, *Shannon* Index and the *Simpson* diversity index. The upper and lower quartiles are indicated by the top and bottom boundaries of the boxplots while the line within the plot indicates the median values. Analysis of variance was performed between patients' variables for each site on each of the diversity measures and was only reported if $P < 0.05$. None of the variables fulfilled this criterion. None of the variables fulfilled this criterion so no significance bars are shown.

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

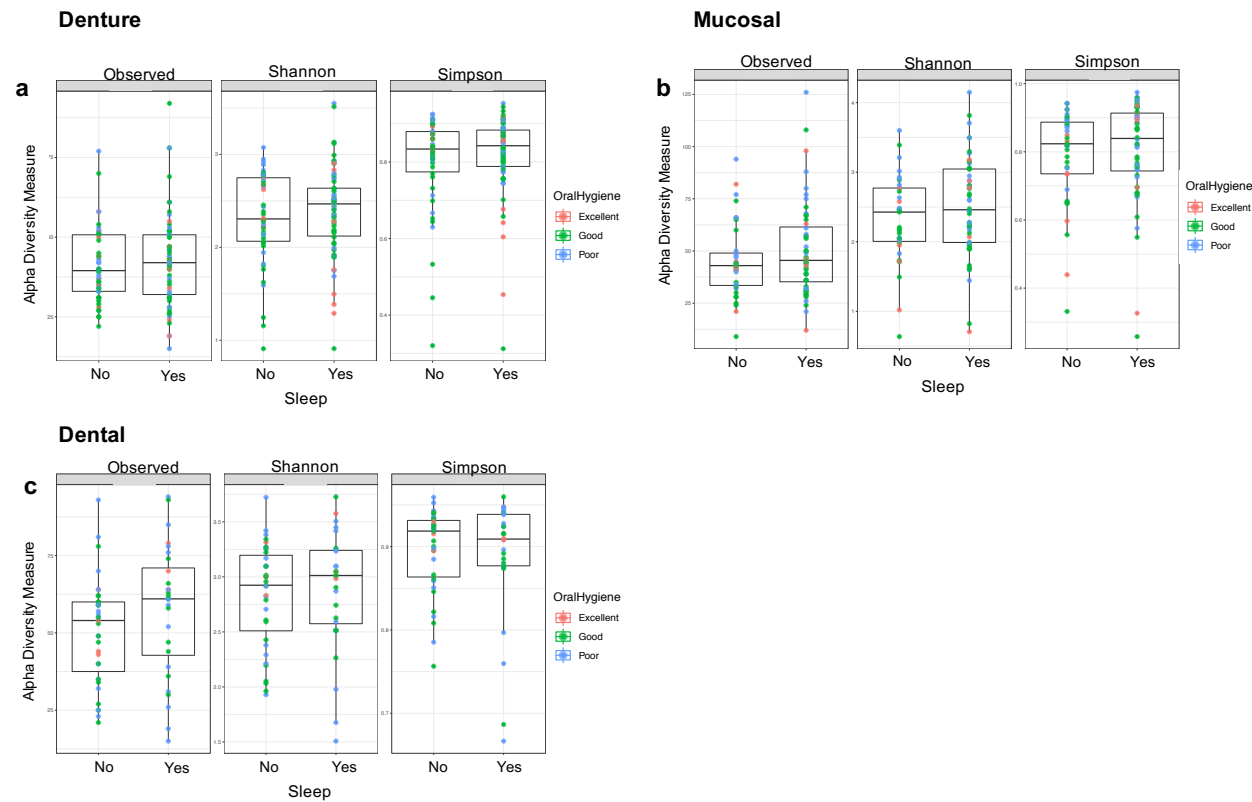


Figure 2.8 **Microbiome diversity measures between sleeping with or not sleeping with dentures *in situ*.** Diversity(observed, Shannon and simpson) between patients that did and did not sleep with their denture *in situ* were measured for each of the three oral sites: denture (a), mucosal (b) and dental (c) using the observed OTUs, *Shannon* Index and the *Simpson* diversity index. The upper and lower quartiles are indicated by the top and bottom boundaries of the boxplots. Analysis of variance was performed between patients' variables for each site on each of the diversity measures and was only reported if $P < 0.05$. None of the variables fulfilled this criterion. None of the variables fulfilled this criterion so no significance bars are shown. while the line within the plot indicates the median values. Analysis of variance was performed between patients' variables for each site on each of the diversity measures and was only reported if $P < 0.05$. None of the variables fulfilled this criterion.

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

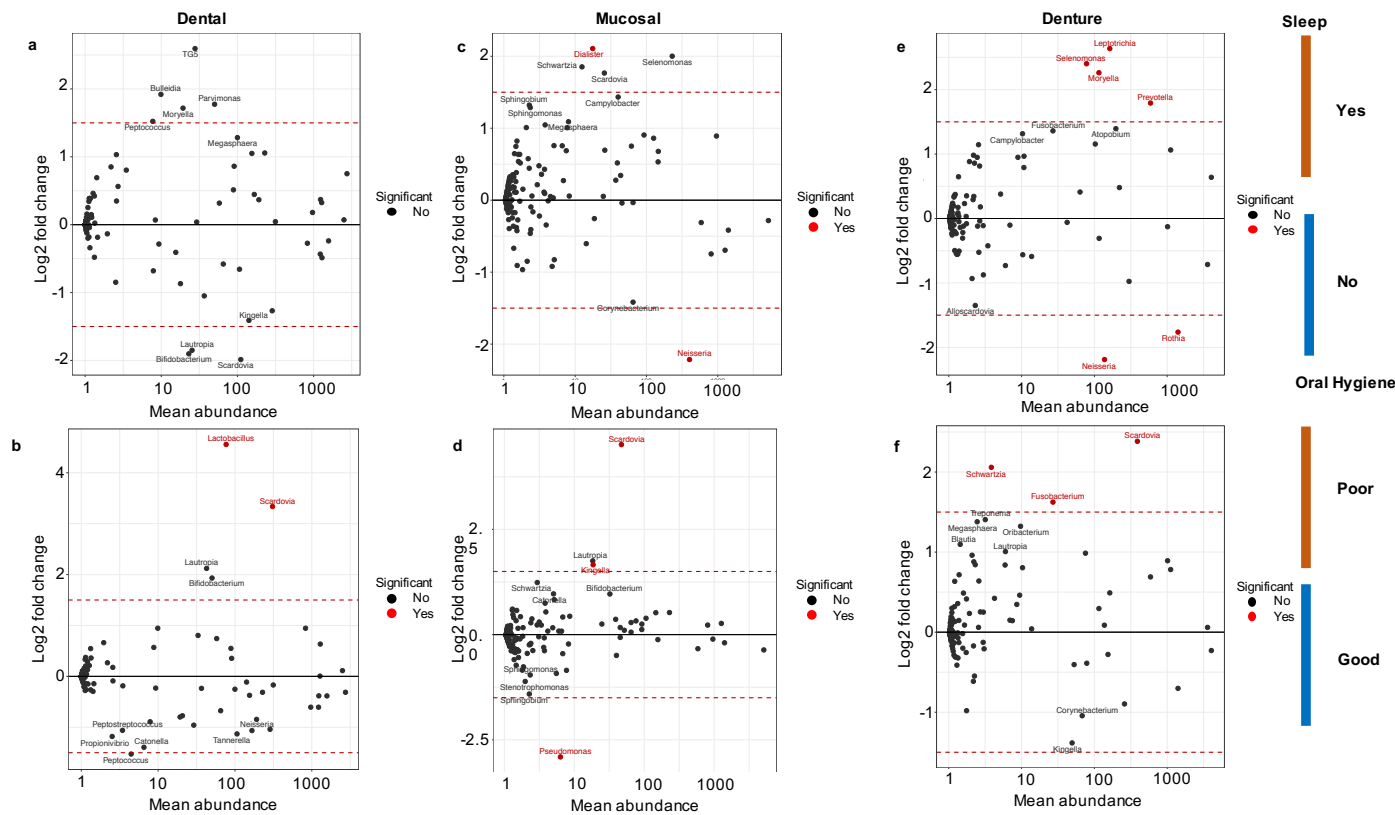


Figure 2.9 **Measure of taxa in significantly higher abundance between patient variables.** MA plots, which are Log2 fold change plotted the against mean abundance, depict taxa that are in differing levels of abundance between patient variables. The patient variables are whether the patient sleeps with their denture in and their overall hygiene rating. The level of each genus between the two patient variables in the Dental (A, B), Mucosal (D, C) and Denture (E, F) microbiomes are indicated by each individual dot. The top ten taxa are labelled with their genus name and significant taxa, with an FDR adjusted p value < 0.01, are indicated in red. Plots A,C and E represent the Log2 fold change of those taxa in higher abundance in those who did sleep (+ve Log2 fold change) in their denture compared to those who didn't (-ve). Plots B,D and F represent taxa that were in higher abundance in those with poor (+ve Log2 fold change) compared to those who had good (-ve) oral hygiene.

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

2.4.4 Influence of *Candida* load on hygiene and microbial communities

The levels of *Candida* were next compared, which were ascertained by qPCR and converted to colony forming equivalent (CFE) counts and normalised to bacterial CFEs using amplification of ITS and 16s (Kraneveld et al., 2012, O'Donnell et al., 2015b). The *Candida* load was then compared between the three denture hygiene metrics denture cleaning frequency, oral hygiene and whether the denture was left *in situ* whilst sleeping.

When comparing the overall *Candida* load between those who did and those who did not sleep in their denture, we observed that there was no discernible difference (Figure 2.10A). This was found to be true at each of the oral microbiome sites denture, mucosal and plaque. We also found that there was not significant difference in the *Candida* load when testing with ANOVA. Similarly, when comparing the *Candida* load between those with poor, good or excellent oral hygiene at each of the three sites there was no significant difference in the overall *Candida* burden (Figure 2.10B). The frequency of denture cleansing appeared to have no visible effect on the *Candida* load on denture, mucosal or plaque samples and there was no statistical significance found when using an ANOVA to compare the two cohorts.

In addition to the impact of oral hygiene we considered the relationship between the levels of *Candida* and the composition and diversity of the bacterial community. Diversity as measured using Bray-Curtis was used to elucidate differences in bacterial diversity due to *Candida* load. *Candida* was compared between low, medium and high loads. From the NMDs there is no distinct separation of clusters relating to the low, medium and high *Candida* loads due to little dissimilarity between the samples (Figure 2.10A). However, a slight trend in the ordination of the points can be observed in relation to the abundance of *Candida*, as shown by the gradient of *Candida* load (Figure 2.10B).

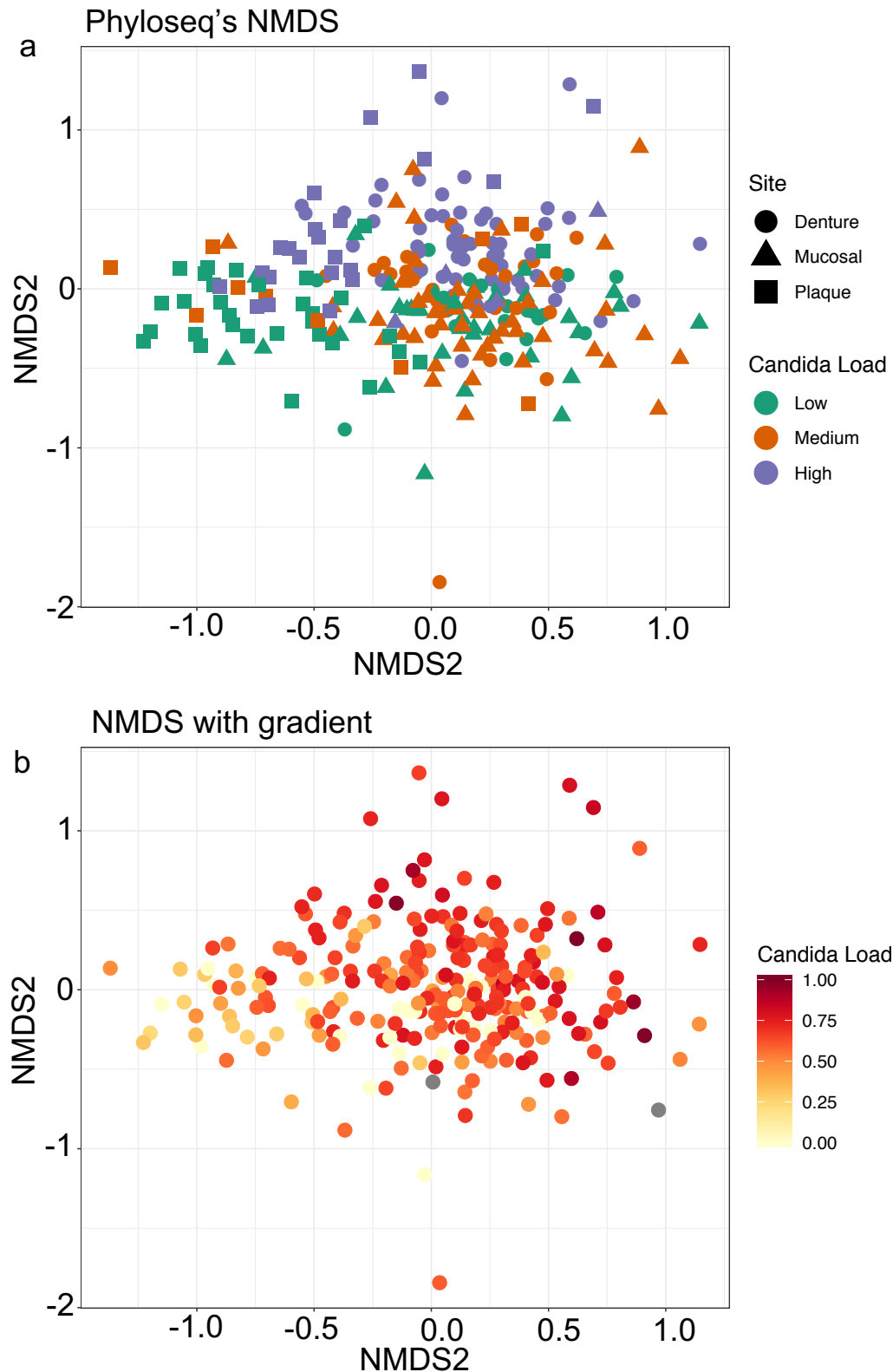


Figure 2.10 **Non-metric dimensional scaling of OTU data based upon the Bray-Curtis distance measure.** NMDS of Bray-Curtis beta diversity metric of oral microbiomes community data is indicated by oral site using shapes and *Candida* load is indicated as either low, medium or high by colour (A), community data is represented as a gradient of *Candida* load which is the total *Candida* CFE/bacterial CFE (ITS/16S) (B).

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

When comparing diversity between the low, medium and high *Candida* loads we observed that the overall abundance of different OTUs slightly decreases. Similarly, when comparing the diversity between low, medium and high *Candida* with the *Shannon* and *Simpson* diversity measures there is a small reduction in diversity, from the low to the medium and high *Candida* load (Figure 2.11). No significance in diversity was found using any of the three measures between low, medium and high *Candida* load.

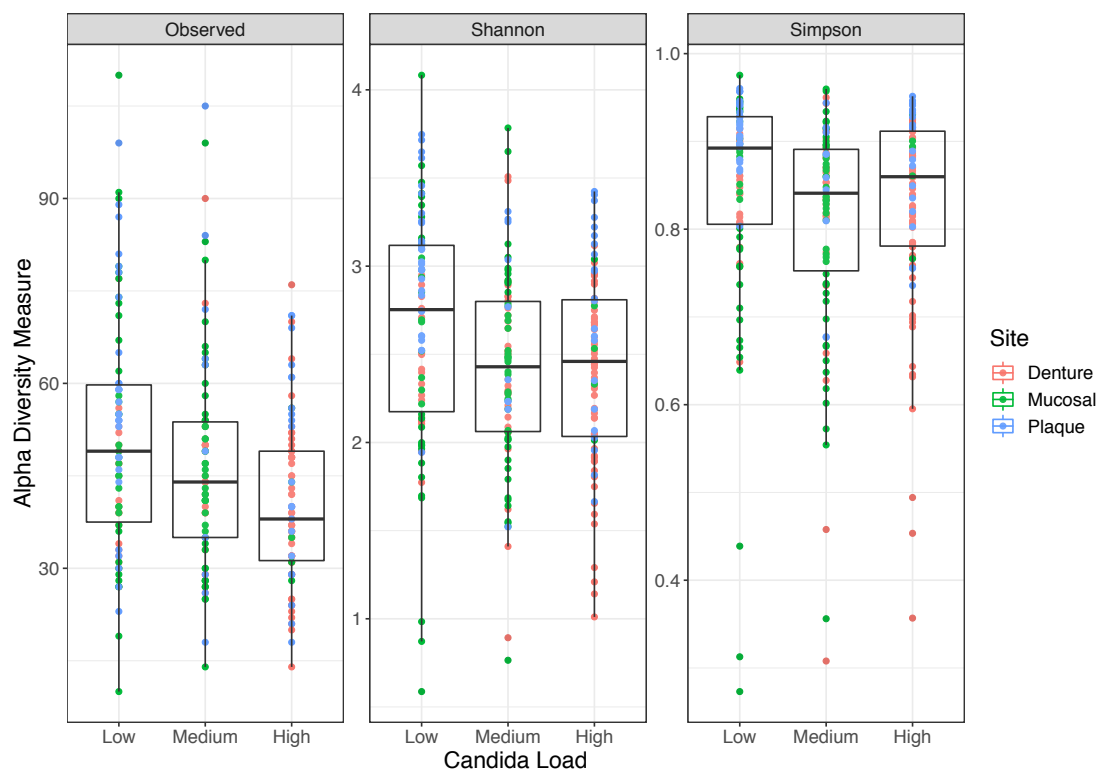


Figure 2.11 **Microbiome diversity measures between *Candida* load low, medium and high.** Diversity by the between patients with low, medium or high *Candida* burden were measured for all patients. The diversity metrics included are the *Observed* number of OTUs, *Shannon* and *Simpson* diversity indexes. The upper and lower quartiles are indicated at the top and bottom boundaries of the boxplots while the line within the plot indicates the median values. Analysis of variance was performed between patients' variables for each site on each of the diversity measures and was only reported if $p < 0.05$. None of the variables fulfilled this criterion. None of the variables fulfilled this criterion so no significance bars are shown.

Finally, hygiene measures including the overall hygiene score, the OHIP score and denture cleaning frequency were all tested for correlations with genus of bacteria. OHIP score and denture cleaning frequency did not correlate significantly with any

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

genera of bacteria, as illustrated in Figure 2.12. The bacterial genus *Scardovia* was positively correlated with oral hygiene, implying an increased level of *Scardovia* within the mucosal microbiome as oral hygiene measures diminishes. A higher *Candida* load was correlated with a significantly higher level of genera including *Lactobacillus* at all three oral sites. Within the mucosal microbiome the abundance of the genera *Acineobacter*, *Faecalibacterium*, *Janthinobacterium*, *Halomonas* and *Shewentalla* is positively correlated with and increased *Candida* load. Within the plaque microbiome *Scardovia* is positively associated with increased levels of *Candida*. Conversely specific genera such as *Leptotrichia* are negatively associated with *Candida* in both the plaque and denture microbiome. Other significant negative correlations include, *Tannerella* (plaque), *Captnocyphaga* (plaque), *Fusobacterium* (denture & mucosal), *Oribacterium* (mucosal) and *Haemophilus* (mucosal). These data suggest that *Candida* species have a subtle influence on the bacterial microbiome in denture patients and can significantly influence specific genera. Oral hygiene measures had less influence on bacterial genera comparatively.

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

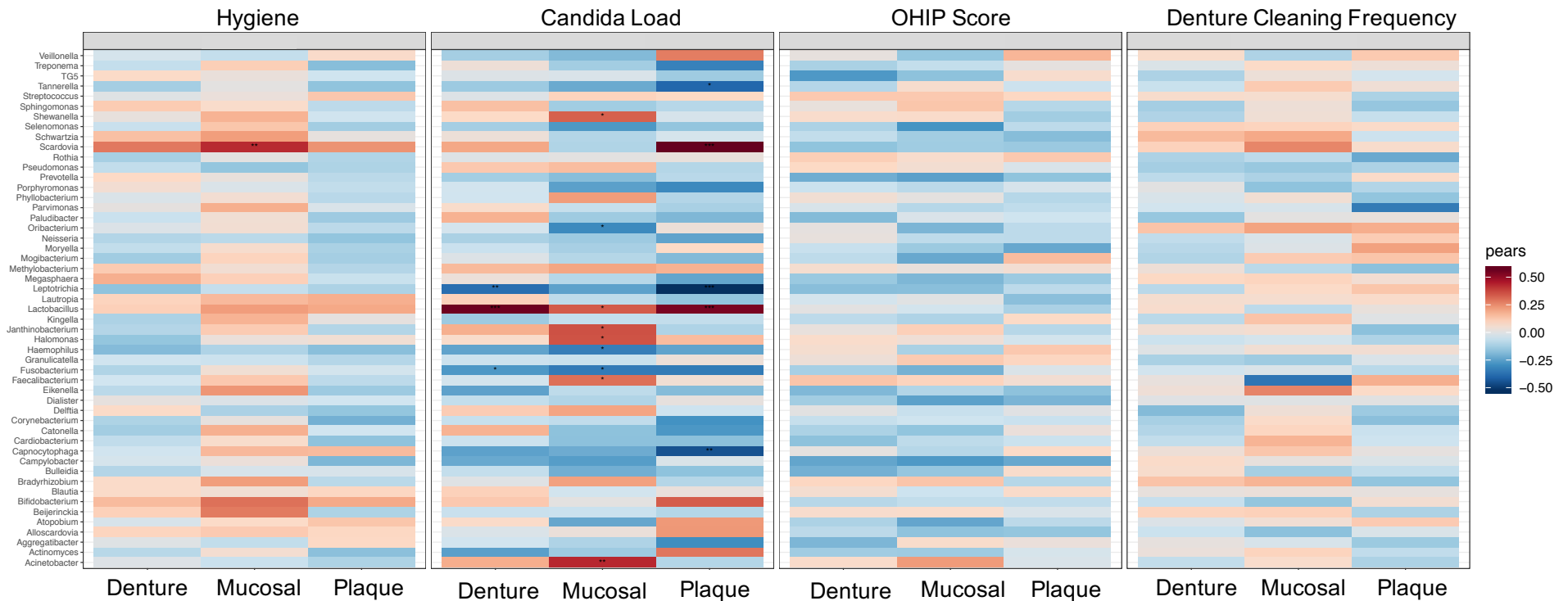


Figure 2.12 **Correlations of variables with the levels of specific Genus on the mucosal, denture and dental surface.** Heatmap depicting the specific correlations between clinically relevant patient meta-data. This meta data is comprised of physician measured (Hygiene and OHIP Score) and patient reported data (Denture Cleaning Frequency) as well as the relative *Candida* load (ITS/16s). Correlations were performed using the Pearson (pears) coefficient and *p*-values were corrected using the Benjamini Hochberg false discovery rate. Correlations are shown between the bacterial genus and the clinical measures for each oral site which was either denture, mucosal or plaque. Significance is indicated by corrected $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***.

2.5 Discussion

As the elderly population expands, the number of denture wearers will coincidentally rise. In the UK population, approximately 20% wear removable dentures, with 70% of UK adults older than 75 years old wearing dentures (Hannah et al., 2017). Dentures can influence oral health status, particularly in relation to the oral microbiome. Other patient related factors may also alter the environment, such as denture hygiene. Poor oral hygiene is often to blame for associated oral inflammation, exacerbated by denture cleanliness, age of denture, salivary flow, diet and smoking (Martori et al., 2014). Soft tissue inflammation results from persistent exposure to microorganisms, a characteristic of DS (O'Donnell et al., 2017). Numerous bacterial species and yeasts frequently adhere to the denture surface and form a biofilm amongst cracks and crevices of acrylic substrates (Hannah et al., 2017). Here it is reported for the first time the relationship between denture hygiene practices, the oral microbiome and yeasts. The data presented demonstrates importance and resilience of *Candida* species in the denture wearing population, and how its presence has a bearing on the bacterial microbiome.

Currently there are a limited number of denture related microbiome studies available in the literature (O'Donnell et al., 2015b, Shi et al., 2016, Yitzhaki et al., 2017). However, we were able to utilise these data to develop a representative model of denture plaque based on the most dominant genera represented (Ramage et al., 2012c), a concept *in vitro* denture model based on an earlier concept (Sherry et al., 2016a). This *in vitro* denture cleansing study revealed that the bacterial component of a 9 species biofilm was significantly more sensitive to chemical and mechanical disruption than the *C. albicans* component. The study revealed that despite clear differentiation from an untreated biofilm, the levels of live *C. albicans* retained on the acrylic surface did not change when challenged. These data suggested that a baseline level of *C. albicans* was retained that was less sensitive to an oral intervention than bacteria. The use of a more sensitive molecular assay was a primary reason for observing this effect, one that would otherwise be missed using standard microbiological plating (Ramage et al., 2012c).

The results from these analyses prompted us to revisit microbiome data we had obtained in a previous microbiome analysis, where primary outcome measures were focussed on disease

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

subtypes (O'Donnell et al., 2015b). In this study design we had collected patient-related data, including oral hygiene information. Given the *in vitro* analysis outcomes, it was hypothesised that *Candida* species present in the clinical samples may also be less affected by oral hygiene interventions, with a caveat to the analysis that this was cross-sectional in nature. These data supported the notion of biofilm insensitivity, but instead of *Candida* being unimpacted alone, the bacterial microbiome was shown to also be uninfluenced by routine oral hygiene practices (composition and diversity). Interestingly though, the specific abundance of individual genera was observed, both on the denture and the mucosa, including genera such as *Leptotrichia*, *Selenomonas*, *Moryella*, *Prevotella* and *Dialaster* in those who slept with their denture *in situ*. Moreover, poor oral hygiene resulted in *Scardovia* and *Lactobacillus* at significantly higher abundance, along with *Fusobacterium* and *Schwartzia* in the denture microbiome. Whether these interactions constitute a synergistic or antagonistic relationship with *Candida* species remains to be determined, but at least suggest subtle dysbiosis correlating with reduced oral hygiene standards.

Next, to establish the impact of *Candida* load on any microbiome changes, and how these were impacted by oral hygiene measures, we normalised both levels based on bacterial load by qPCR according to established methods (Kraneveld et al., 2012). Denture cleansing frequency and the other measures appeared to have no visible effect on the *Candida* load on denture, mucosal or plaque samples, a result mirroring our own *in vitro* observation. Indeed, there are no measurable changes in diversity indices across the different *Candida* loads. One of the caveats of the study design is the cross-sectional nature and lack of power, thus non-significant results between these variables are not necessarily absence of effect, but rather a result of not achieving the optimum sample size required. Interestingly, although the bacterial microbiomes were not significantly influenced by *Candida* load, a slight trend can be observed in relation to the abundance. This suggests again that subtle changes to the microbial composition are reflected by changes in abundance of *Candida* rather than the oral hygiene intervention.

Our final analysis involved correlation analysis at the genus level, looking at a range of parameters, including overall hygiene score, the OHIP score and denture cleaning frequency. This approach enabled us to observe clear positive and negative associations with different

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

oral sites, including the denture surface, and pick out significant positive correlations. This was deemed this an important tactic, as the breadth and depth of literature is now beginning to demonstrate important interkingdom relationships (O'Donnell et al., 2015a). Neither OHIP score and denture cleaning frequency were shown to correlate significantly with any genera of bacteria, though *Candida* load and oral hygiene did reveal significant associations. As has been described elsewhere, a higher *Candida* load correlated with a significantly higher level of genera including *Lactobacillus*. Moreover, the bacterial genus *Scardovia* was positively correlated with oral hygiene. Conversely specific genera such as *Leptotrichia* and *Fusobacterium* are negatively associated with *Candida* on the denture microbiome. Given that it is now understood that *C. albicans* specifically interacts with *Staphylococcus aureus* and oral streptococci using agglutinin-like sequence adhesins (ALS3) (Peters and Noverr, 2013, Kean et al., 2017), and *P. gingivalis* using InlJ, an internalin protein family, to interact with the same ALS3 adhesin (Sztukowska et al., 2018a), then it is unsurprising that we are able to tease out specific interactions. These analytical approaches, while not hypothesis driven *per se*, will help pinpoint the bacterial genera we should consider when designing and developing new biofilm models of microbial pathogenesis. Moreover, understanding the important elements of polymicrobial interkingdom interaction, no matter how subtle, could provide useful direction in the development of novel antibiofilm strategies. The concept of the mycofilm (Kean et al., 2017), where bacteria utilise bacteria as a scaffold to support their own biofilm, is a prime reason we ought to consider *C. albicans* as the real keystone oral microorganism (Janus et al., 2016). Targeting this may be crucial in generating a wider anti-biofilm effect.

Although we did not observe any consistent effects of hygiene on the microbiome within this work poor oral hygiene has been previously associated with a worse case of DS. Studies have previously highlighted that there is a tendency within the denture wearing population for poor maintenance of their dentures. Additionally it has been previously shown that there is a relatively low percentage (11.9% in one study) and that poor denture hygiene was associated with an increased incidence of DS (Dikbas et al., 2006). Differences in denture cleaning methods and sleeping habits have also been attributed to differences in the incidence of DS (Martori et al., 2014, Sadig, 2010). The prolonged isolation of the denture from the beneficiary effects of the normal salivary flow are hypothesised to be responsible for this effect. The microbial community under the denture is sealed and isolated from the rest of the oral

Chapter 2: The impact of *Candida* and its interactions with the microbiota at the denture surface

environment allowing a pathogenic community to exist and thrive on the denture surface (Sobel et al., 1992).

Microbiome analysis is relatively new in the context of the study of microbes and microbial communities. At the time of sequencing for this study the read depths were more limited with the read depth for some samples at <1000 bases; current platforms can generate >30,000 reads per sample. Additionally, rarefaction was used within this study as a method of determining the minimum acceptable sample depth and, in turn, subsampling all samples down to said minimum read depth. With the current data an alternative method would be to retain the entire resolution of the data with no rarefaction and a repeat study would more than likely yield much higher levels of reads and hence improve the inferred taxonomic composition. In addition to this, at the time of study conception and design ITS sequencing for identification of fungal species was not widely adopted, although had been shown to speciate fungi, and therefore the cruder identification of fungal load by qPCR was utilised. The mycobiome is still under-utilised and is a weakness within this study and in future studies. As we have previously discussed, mycobiome studies are slowly on the rise and any future community profiling within the oral cavity would benefit from having joint ITS/16s sequencing such as in the case of recent studies carried out in dental caries (O'Connell et al., 2020).

In summary, this study has been the first to specifically investigate the relationship between denture hygiene, the oral microbiome, and the influence of *Candida* species in denture wearers. The findings from this study suggest that maintaining good denture hygiene and hygiene practices do not appear to have a strong influence on altering the microbiome but taken positively this indicates a stabile microbial population. *Candida* species appear to be more resilient to the daily treatments which was mirrored in patient demographics, where denture cleaning and hygiene does not impact the fungal load. We predict this resilience is a key factor in *Candida's* role within the oral microbiome. Therefore, future studies in oral microbiology and beyond should pay closer consideration to the mycobiome and the influence it can have on the bacterial microbiome. However, to better understand this it is important to understand the driving mechanisms that underpin biofilm formation in *C. albicans*.

2.6 Highlights

- *Candida* is less sensitive to a daily denture treatment regimen than oral bacteria species. The fungal load is not impacted by oral cleansing or oral hygiene measures indicating its high resilience and stability in the oral microbiome.
- *Candida* load has an influence the overall denture and oral microbiome with shifts in the beta diversity being observed in response to fungi.
- Hygiene measures have limited impact on the microbiome of individuals with few bacterial species being significantly different between good and poor oral hygiene. The diversity of the oral microbiome was not significantly affected by any of our oral hygiene or cleanliness measures.
- Contrary to common assumptions correlations of co-existence of *Candida* and the *Lactobacillus* species in oral biofilms exist. *Lactobacillus* was positively correlated with $p > 0.05$ in the mucosa, dentate and edentate surfaces within the oral cavity.

3 Transcriptomic profiling of phenotypically distinct *Candida albicans* clinical isolates

3.1 Background and Introduction

Biofilms are a major clinical concern due to their persistence, antimicrobial resistance, host evasion and other tolerance mechanisms (Ramage et al., 2012b). *Candida* biofilms are typically defined as an aggregate of cells attached to one another and/or a surface and encapsulated within an extracellular matrix (Nobile and Johnson, 2015). It is this aggregation and encapsulation of the cells that confers microbial tolerance. Biofilm formation is an important factor in *C. albicans* pathogenesis, however, there is variability from strain to strain in biofilm forming capability (Tumbarello et al., 2007, Tumbarello et al., 2012). There are various environmental and physiological factors, including nutrient availability, interaction with host/bacterial species, hypoxia, pH, and temperature that influences the phenotype of *Candida* species. This variability in biofilm formation has been categorised *in vitro* from clinically derived isolates. Within a population *C. albicans* isolates can range from low biofilm forming (LBF [yeast-like poorly aggregated strains]) to high biofilm forming (HBF [hyphal and extremely aggregated in their phenotype]), under the same conditions. This was shown to be true in clinically derived samples of patients with candidemia. The biofilm phenotype within this cohort was associated with increased tolerance to antimicrobials and increased pathogenicity (Sherry et al., 2014). This variability is not unique to bloodstream isolates but is also observed within other sites such oral and vaginal infections (O'Donnell et al., 2017, Sherry et al., 2017).

Previously, the role of different metabolic pathways including amino acid metabolism has been determined in *C. albicans* biofilm heterogeneity (Rajendran et al., 2016b). Our understanding of *C. albicans* adaptation to the host environment have to date been mostly achieved from studies concerned with planktonic cultures. For the first time this project aimed to dissect the role of adaptive phenotypes of *C. albicans* biofilms in response to nutrient stimuli/stress. The control of cellular physiology in response to immediate environmental changes involves reprogramming of gene expression regulatory systems. A distinct component of biofilm adaption in *C. albicans* is phenotypic switching and growth, which is stimulated by a variety of factors, including changes to the carbohydrate source and amino acid starvation (Tripathi et al., 2002, Brown et al., 2014a, Sudbery, 2011). Changes in

functional and metabolic pathways impact *C. albicans* pathogenicity, which includes yeast-hypha morphogenesis, phenotypic switching, adhesins, invasins, and secreted hydrolases (Brown et al., 2014a), factors all important to the biofilm phenotype (Nobile and Johnson, 2015). Notably, biofilm related processes within *C. albicans* are driven by 6 key transcriptional regulators as part of a complex transcriptional circuitry (Nobile et al., 2012). *C. albicans* biofilm formation therefore seems to depend upon two factors: a genetic predisposition for enhanced biofilm formation and the correct external stimulus to induce the phenotype.

The use of RNA-Seq, i.e., the identification of expressed mRNA transcripts through using high throughput sequencing technologies such as Illumina, has become common and, in some instances, routine within many medical and life science disciplines. The same is becoming true of microbiological research and the identification of transcriptional changes in *Candida* spp. (Chong et al., 2018b). Transcriptomics holistic identification of gene expression, this allows for identification of functional groups and network/pathway regulatory processes. Similarly, within fungal and *Candida* infection studies, RNA-Seq is a demonstrably powerful technology in discerning perturbations in the cellular response of *C. albicans* to different stimuli and environmental conditions as well as antimicrobial treatment responses (Romo et al., 2019, Burgain et al., 2019, Burgain et al., 2020). RNA sequencing has become a popular tool for discerning transcriptional changes, offering higher resolution compared to qPCR and microarray technologies at an ever-decreasing cost. This, in combination with the increase in bioinformatic pipelines, focussed upon discerning changes within networks and pathways allow for enhanced interpretation of *Candida* transcriptomic datasets.

3.2 Aims

It was hypothesised that molecular profiles could be discerned to differentiate LBF and HBF capacity. This chapter aims to interrogate the different gene expression and regulatory networks involved in biofilm formation. Additionally, it will build on our groups previous findings that metabolic adaptation is key to biofilm formation in *C. albicans*. The following work will interrogate clinical isolates identified as either low biofilm forming (LBF) or high biofilm forming (HBF) isolates both in the presence and absence of a biofilm inducing stimulus (serum).

Specifically, this chapter aims to identify:

- Differences in morphology and phenotype in HBF and LBF in both the presence and absence of serum.
- Evidence of transcriptomic differences between HBF and LBF isolates both in the presence and absence of serum.
- Functionally relevant pathways and networks responsible or related to both these phenotypic differences.
- Transcriptomic and functional pathways/networks related to the biofilm induced phenotype.

This work has been presented at the following conferences:

Eurobiofilms, Amsterdam 2017 - **Integration of Metabolomics and Transcriptomics in Biofilm Research (Poster, Recipient of a travel grant to attend from the organisers)**

ECCMID, Madrid 2018- ***De novo* transcriptome assembly and annotation for analysis of the emerging pathogen *Candida auris* (Poster)**

OMIG, Gregynog 2018- **Omics approaches to studying microbial biofilms in oral health (Talk, Recipient of a travel grant to attend from the organisers)**

IADR, London 2018- **Omics and bioinformatics approaches to studying oral biofilms (Poster)**

Eurobiofilms, Glasgow 2019 - **Transcriptomic and multi-omics data integration approaches to interrogate mono and interkingdom species *Candida* biofilms (Talk)**

Invited talk, Newcastle 2020 - **OMIC modelling of *Candida albicans* interkingdom biofilm interactions (Seminar)**

3.3 Methods

3.3.1 Culture conditions and standardisation

This chapter utilised clinical *C. albicans* (n = 10) bloodstream isolates, collected under the approval of the NHS Scotland Caldicott Guardian's (Sherry et al., 2014, Rajendran et al., 2015). Isolates had been previously characterised in detail and classified as to their biofilm forming ability. Isolates were stored in Microbank® vials (Pro-Lab Diagnostics, Cheshire, UK) at -80°C until sub-cultured onto Sabouraud's dextrose agar (SAB [Sigma-Aldrich, Dorset, UK]). *C. albicans* isolates were propagated in yeast peptone dextrose (YPD) medium (Sigma-Aldrich,

Dorset, UK), cultured for ~16h in a shaking incubator at 30°C, washed by centrifugation and resuspension in PBS twice before being standardised into RPMI-1640 (Sigma-Aldrich, Dorset, UK) to 1×10^6 cells/mL, as described previously (Ramage et al., 2001b).

3.3.2 Biofilm assay

C. albicans biofilms were grown according to our established protocols (Rajendran et al., 2015). Pre-characterised *C. albicans* isolates with high (HBF [n=5]) or low (LBF [n=5]) biofilm forming ability were used throughout this chapter (Sherry et al., 2014, Rajendran et al., 2015). For all experiments, biofilms were grown in polystyrene plates, 75 cm² tissue culture flasks or Thermanox coverslips in RPMI (Sigma) for 90min, 4h or 24h at 37°C. For induction experiments, RPMI media was supplemented with foetal calf serum (FCS [25% (v/v)]) or dialysed serum (DS [25% (v/v)]). Small molecular components were removed from the serum using dialysis membrane tubing. The biofilm biomass of each isolate was assessed with the crystal violet (CV) assay as previously described (Rajendran et al., 2015). Before being quantified using the FLUOstar Omega plate reader (BMG, UK) at 570nm. Hyphal formation was quantified using a light microscope and eye piece graticule and stage micrometre. Measurements were taken for total hyphal length at 90 and recorded before comparisons between HBF and LBF after 90min growth.

3.3.3 RNA extraction and sequencing

Two isolates were selected 1 LBF (204) and 1 HBF (39) based on their biofilm forming capabilities and their inclusion in previous group studies. These isolates had previously been compared a 2016 study (Rajendran et al., 2016b). Following incubation, biofilms were washed with phosphate buffered saline (PBS) to remove any non-adherent cells before a cell scraper was used to dislodge the biomass into 1ml of TRIzol™ (Life Technologies, Paisley, UK). The TRIzol containing the scraped biomass was transferred to a bead beating tube containing 0.5mm sterile glass beads. The biomass then disrupted using a Mini-beadbeater (Biospec Products) for 3 x 30seconds. Following bead beating 100µl of bromo-chloropropane was added to the sample before the sample was vortexed and centrifuged for 15mins at 13000rpm at 4°C. Subsequent to centrifugation samples split into 3 layers, the upper aqueous layer was decanted by pipette into a nuclease free Eppendorf tube. To the Eppendorf an equal amount of ice-cold isopropanol was added ~500µl before being inverted 3-4 times to mix thoroughly and then stored in a - 20°C freezer overnight. The samples were then centrifuged

for 10mins at 13000rpm at 4°C to pellet the RNA and the entire supernatant was removed. The pellet was then washed with 700µl of ice-cold 70% ethanol before being centrifuged for 5mins at 6500rpm at 4°C. The ethanol was then removed, and the pellet allowed to air dry until no ethanol remained. 20µl of RNase free water was then added to resuspend the RNA pellet. Total RNA was DNase (Qiagen, Crawley, UK) treated and purified using the RNeasy MiniElute clean up kit (Qiagen, Crawley, UK), as per manufacturer's instructions. RNA was quantified and quality assessed using a NanoDrop spectrophotometer (ND-1000, Thermo Scientific, Loughborough, UK). Each isolate was grown in triplicate and a minimum of 1 µg of total RNA was submitted for each sample and sent for sequencing to The GenePool (Edinburgh, UK). RNA integrity was assessed using a Bioanalyzer where an RNA integrity number (RIN) value >7.0 was deemed acceptable for RNA-Seq. TruSeq (Illumina v3 i7) stranded mRNA library prep kit was used to prepare libraries before sequencing on an Illumina HiSeq v4 yielding 50 base single end sequences. RNA-Sequencing is summarised in Figure 3.1.

Table 3.1 **Candidemia isolates used in each individual experimental design.** *Candida albicans* isolates from blood stream infections categorised as either HBF or LBF by Ranjith Rajendran. + indicates that the isolates were used in each of the three study designs. Transcriptome profiling, metabolomic profiling or cell wall proteomics. Optical density from the original classification by crystal violet also indicated (OD).

Isolate Number	Biofilm classification	OD (595 nm)	Source	Transcriptomics	Metabolomics	Proteomics
39	HBF	0.861	BSI	+	+	+
48G	HBF	0.542	BSI	-	+	+
177A	HBF	0.728	BSI	-	+	+
17B	LBF	0.015	BSI	-	+	+
140	LBF	0.091	BSI	-	+	+
204	LBF	0.1	BSI	+	+	+

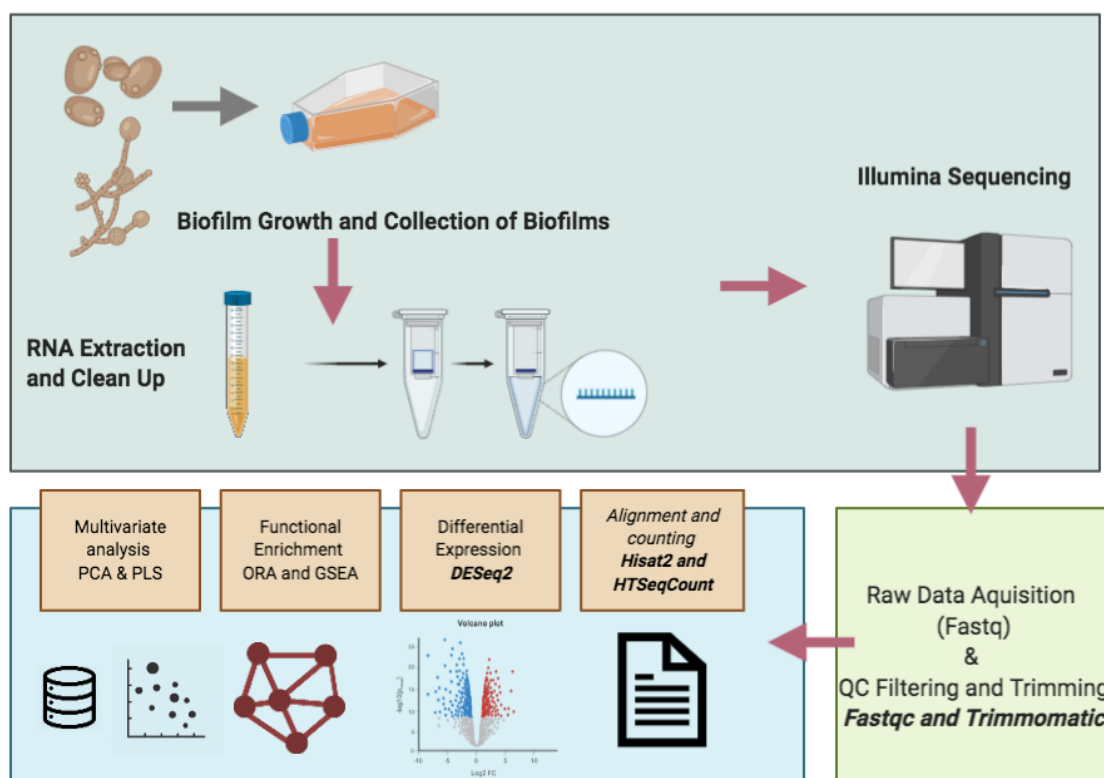


Figure 3.1 **Overview of sample preparation, RNA-Sequencing, and analysis approaches.** Schematic depicting methodology starting with growth of *C. albicans* biofilms from high and low biofilm formers, extraction of RNA by trizol and column clean-up followed by 50bp sequencing, data retrieved from Edinburgh Genomics was then QC followed by differential expression and pathway analysis.

3.3.4 RNA-Seq analysis

RNA-seq reads were processed by first quality controlled using the software Trimmomatic v0.38 (Bolger et al., 2014) to remove Illumina adapters low quality bases leading=3 and trailing=3 and reads with remaining length of less than 30 bases. An index of the *Candida* genome (CGD) database reference *C. albicans* genome (SC5313_A22) was then constructed using Hisat2 (Kim et al., 2019). The current genome maintained by the CGD is diploid, we utilised a haploid genome for RNA-Seq analysis by disregarding the B variants of the chromosomes from the fasta and gff annotation files.

Hisat2 (v2.1.0) was then utilised to map the trimmed sample reads to the SC5314 genome (Kim et al., 2019). Subsequent SAM files containing the aligned reads were then coordinate sorted and converted to BAM format using Samtools (v1.7) (Li et al., 2009). Quality of the alignments was assessed using the software Qualimap (v2.2.2) and the BAM files for each

sample were counted to obtain gene counts using the corresponding SC5314_A22 gff file from the *Candida* genome database with the use of the program HTSeq-count (0.11.0) (Anders et al., 2015, Okonechnikov et al., 2016). All gene counts were then parsed into a single large array containing all the of the samples and the corresponding counts for each of the genes. The gene count array was then analysed within the R programming environment assisted by the R Studio GUI (<http://www.rstudio.com/>). Differential expression analysis was largely performed with the assistance of the DESeq2 (v1.26) R package (Love et al., 2014). DESeq2 uses a negative binomial model to estimate gene abundance and differentia expression between variables. Differential expression was performed in a pairwise fashion between sample variables and significance was determined if genes had a Log2FC >1.5 and an FDR adjusted p-value of <0.05. In house scripts were written to perform analysis and visualisations. These scripts were based around many of the highly used and well validated R packages. These included Ordination plots MDS and PCA were performed using DESeq2 and the package pcaExplorer (v2.12.0) (Marini and Binder, 2019). Volcano plots were made with the R package EnhancedVolcano (v1.4.0) (Blighe K, 2020). Additional plots were also drawn with the use of the R package ggplot2 (v3.3.0) (Wickham, 2009). *C. albicans* transcripts that were significantly differentially expressed between pairwise comparisons underwent further functional analysis within the network software Cytoscape (v3.7.2) (Shannon et al., 2003). The plugin GlueGO annotated and grouped the genes into functional categories and significantly over-represented categories were found using the hypergeometric test which deemed functional categories to be enriched with an adjusted p-value <0.05. Networks of over-represented functional categories using either the gene ontology or the KEGG databases were constructed using GlueGO (v2.5.5) plugin and drawn within the Cytoscape environment (Bindea et al., 2009). Additionally, fgsea (v1.12.0) was used to perform GSEA from within the clusterProfiler (v3.14.3) R package (Yu et al., 2012). Multivariate analysis and important feature identification were performed through the mixOmics (v6.10.9) R package utilising their sPLS and PLS-DA functions (Rohart et al., 2017).

3.3.5 Electron microscopy

Two *C. albicans* clinical isolates, utilised in transcriptomics experiments, HBF (39) and LBF (204) were grown on Thermanox™ in the presence or absence of serum for 24h (Table 3.1). After incubation period, biofilms were carefully washed with PBS and then fixed with a fixative

solution containing 2% paraformaldehyde, 2% glutaraldehyde and 0.15 M sodium cacodylate, and 0.15% w/v alcian blue (pH 7.4) and left in solution for 18h. The fixative was discarded, and the biofilms were then covered with 0.15M sodium cacodylate buffer and stored at 4°C until processing (Erlandsen et al., 2004). Briefly the samples were washed 3 × 5min with 0.15 M cacodylate to remove the glutaraldehyde. Following this samples were treated using a 1% osmium tetroxide solution containing 0.15 M sodium cacodylate (1:1). Samples were then incubated in the fume hood for 1h. Samples were rinsed 3 × 10min with distilled water and then treated with 0.5% uranyl acetate and incubated in the dark for 1h. Uranyl acetate was removed from the samples and quickly rinsed with water before a series of dehydration steps were carried out. The biofilms were then washed twice for 5 min each of 500µL 30, 50, 70 and 90% ethanol this was then followed by 4 × 10min washes of 500µL absolute followed again by 4 × 10min washes with 500µL molecular sieved absolute alcohol. Hexamethyldisilazane (HMDS) was used to dry the specimens by soaking the samples for 5min before transferring to a plate containing fresh HMDS. All samples were then placed in a desiccator overnight to allow evaporation of any residue and drying. The specimens were then mounted and sputter-coated with gold in an argon filled chamber, and then viewed under a JEOL JSM-6400 scanning electron microscope.

3.3.6 Cell wall proteomic analysis

Cell wall processing was performed by the postdoctoral researcher of the University of Glasgow Dr. Ranjith Rajendran. Furthermore, the processing and acquiring of data was performed by Dr. Ranjith Rajendran at the University of Aberdeen under the supervision of Prof. Carol Munro.

Cell walls were extracted as described previously (Mora-Montes et al., 2007). Briefly, *C. albicans*, (HBF [n=3]) or low (LBF [n=3]) as described in table 3.1, biofilms grown in the presence and absence of FCS for 24h which were then dislodged and centrifuged at 3,000 × g for 5min, washed once with chilled deionized water, resuspended in deionized water, and mechanically homogenised with glass beads in a FastPrep machine (Qbiogene, Fisher Scientific, UK). The disrupted cells were collected and centrifuged at 5,000 × g for 5min. The pellet, containing the cell debris and walls, was washed five times with 1 M NaCl, resuspended in buffer (500 mM Tris-HCl buffer, pH 7.5, 2% [wt/vol] SDS, 0.3 M β-mercaptoethanol, and 1 mM EDTA), boiled at 100°C for 10min, and freeze-dried.

Proteomics analysis was performed on the freeze-dried cell wall fractions by Aberdeen Proteomics at the University of Aberdeen. The pellets were digested with trypsin according to the PRIME-XS protocol. Mass spectrometry analysis was performed using a Q-Exactive Plus (Thermo Fisher Scientific) and tryptic peptides were identified using the MASCOT searching engine (Matrix Science). Data was processed using the Proteome discoverer v1.4 software quantifying and identifying proteins with >2 peptides against protein database derived from the *Candida* Genome Database (CGD) <http://www.candidagenome.org/>.

3.4 Results

3.4.1 Assessing the Biofilm Phenotype

Initially the biofilm phenotype was assessed comparing LBF/HBF isolates in the presence (+ FCS) and absence (-FCS) of FCS. Addition of FCS was shown to significantly induce biofilm formation of LBF isolates by 5 to 7.4 times compared to isolates grown in RPMI control ($p < 0.01$) (Figure. 3.2A). At 90min LBF+FCS samples showed more hyphal cells with greater extracellular matrix compared to predominantly yeast cells in control LBF-FCS samples. At 24h matured biofilm phase, LBF+FCS formed multi-layered hyphal structures with the appearance of being 'glued' together with extracellular matrix (Figure. 3.2B). HBF isolates have an increased biomass compared to LBF in RPMI and this does not significantly change with supplementation of serum (Figure 3.2A). LBF on the hand have a more comparable biomass to the HBF in the presence of serum. These observations seem to consistently find that the supplementation of LBF with FCS produces many of the features of HBF.

Both the serum supplemented and non-supplemented HBF showed similar phenotypes regarding hyphal and biofilm formation, however, there was observably larger amounts of material attached to the hyphae (Figure 3.2B).

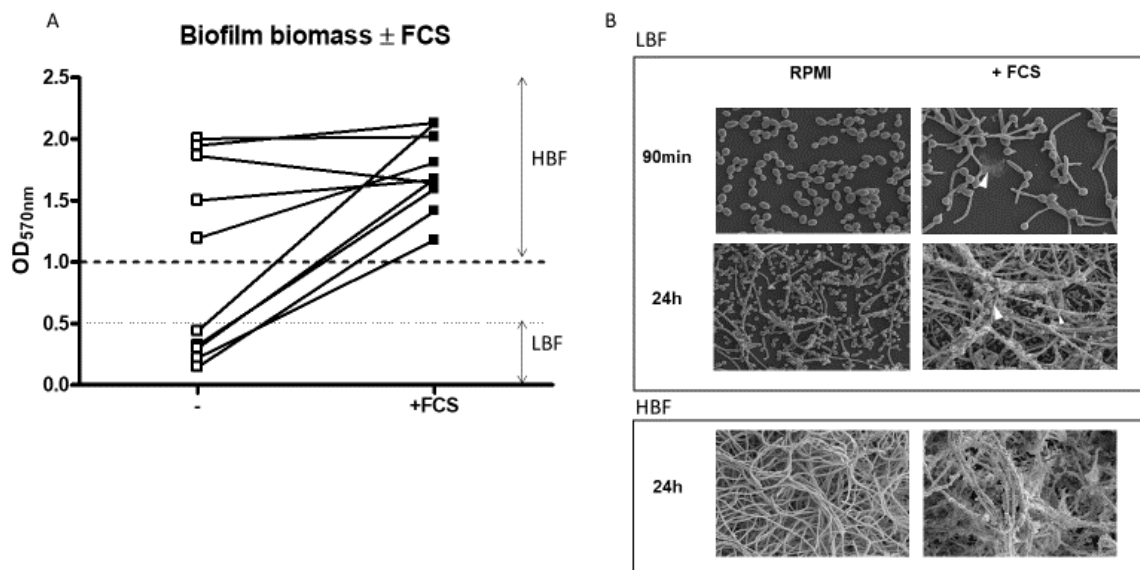
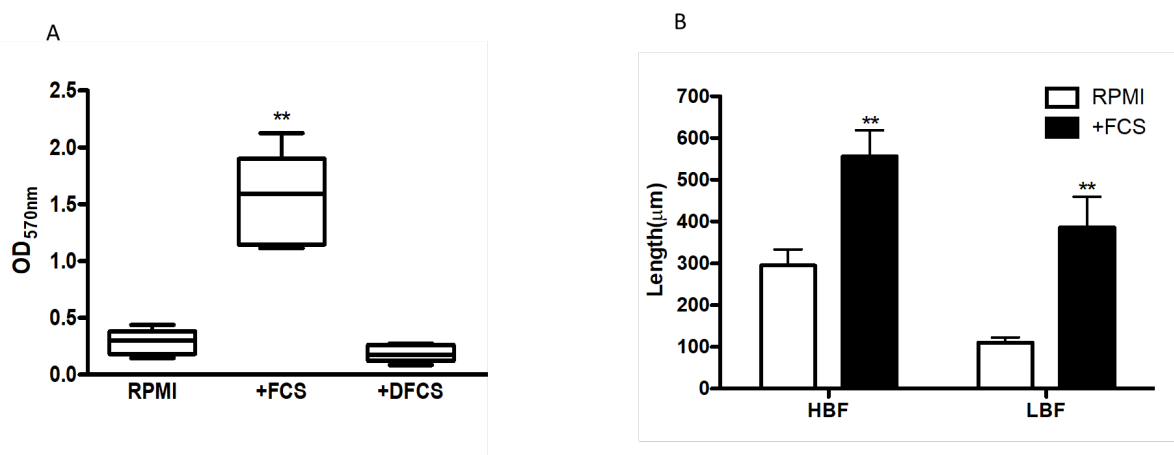


Figure 3.2 **Variation in *C. albicans* biofilm formation in the presence and absence of FCS.** Biofilm formation after 24h assessed using crystal violet assay of 5 low biofilm forming (LBF) and 5 high biofilm forming (HBF) isolates grown in RPMI or RPMI supplemented with serum (A). Scanning electron microscopy (SEM) images of high biofilm forming (HBF) or low biofilm forming (LBF) isolates morphology grown similarly in the presence and absence of serum grown on coverslips for 90min or 24h (B).

Overall, the biomass assays show an increase in biofilm formation in the presence of serum. This is consistent with an observation of a shift in cell morphology, increased hyphal formation and extracellular material.

Additionally, germ tube formation was measurably different at 90min in both HBF and LBF *Candida* isolates. Germ tube formation was shown to be 3.5 and 1.8-fold higher in the presence of serum at 90mins in LBF and HBF respectively (Figure 3.3B). These observations seem to consistently find that the features associated with biofilm formation in *C. albicans* are apparent in the non-biofilm formers when they are grown in the presence of serum.

Figure 3.3 Biomass of LBF isolates grown in RPMI and RPMI supplemented with 25% serum.



The effect of dialysis on the serum prior to supplementation was tested with biofilms grown in RPMI, RPMI supplemented with 25% serum or 25% dialysed serum for 24h. Significance was measured compared to RPMI control to RPMI supplemented with serum or dialysed serum by ANOVA with multiple comparison test showing significant differences ($p < 0.01^{**}$) between RPMI and serum (A). Bar chart depicting the overall length of hyphae in LBF and HBF candida isolates in the media RPMI and RPMI supplemented with 25% serum after 90min. Germ tube length measurements were taken and compared between groups using ANOVA with multiple comparison test displaying significant differences between germ tube length in RPMI compared to RPMI supplemented with serum in HBF and LBF ($p < 0.01^{**}$) (B).

3.4.2 RNA-Seq read quality

The HiSeq platform from Illumina provided an average of 24.4 million reads per sample. With the lowest sample (90m_S39_3) having 13.6 million reads and the highest sample having 52.4 million reads (4h_S39_2) following Trimmomatic quality and adapter removal. The average GC of all reads was 39% and the average duplication rate per sample was 83%. All samples were deemed to have sufficient sequencing depth, greater than >10 million reads per sample, and were processed with Hisat2 with a 95% average alignment rate to the reference genome. These alignments were then assigned to a feature within the *C. albicans* (v22) gene feature format (GFF/.gff) file and the number of assignments counted by the software HTSeq-counts. Of the aligned reads an average of 75% of them were assigned a gene feature. These feature counts were subsequently analysed between samples for differential gene analysis. All sequencing statistics, alignment and assignments are summarised in Table 3.2.

Chapter 3: Transcriptomic profiling of phenotypically distinct *Candida albicans* isolates

Table 3.2 RNA-Sequencing, quality control and alignment statistics. Table depicting the total number of sequences for each sample before trimming, total following trimming and the percentage of duplicate reads and the GC percentage of all reads per sample. Hisat2 was used to align reads and the total number of reads aligned to the reference genome per read are shown with number of reads aligned once, multiple times or not at all shown. HTSeq was utilised to count the number of aligned reads that corresponded to a feature (exon) the percentage and total number of identified features per sample are shown.

Sample ID	Hisat 2				HTSEQ Count		Sequence QC			
	Percentage Aligned (%)	Aligned Once (total read number)	Not Aligned (total read number)	Aligned Multi (total reads)	Percentage assigned a gene feature (%)	Number of reads with a feature (million)	Number of sequences before QC (million)	Percentage duplicate reads post QC (%)	GC percentage of reads post QC (%)	Total number of reads post QC (million)
24h_R204_1	94.49	29179312	1775628	1278384	75.30%	25.4	32.3	82.80%	38%	32.2
24h_R204_2	94.39	24990992	1550537	1104150	75.60%	21.9	27.7	81.50%	38%	27.6
24h_R204_3	94.51	36981739	2248905	1741642	75.10%	32.4	41	84.40%	38%	41
24h_R39_1	94.09	21449725	1418447	1120784	70.20%	17.7	24.1	84.80%	39%	24
24h_R39_2	96.12	15927716	681417	931221	72.40%	13.4	17.8	84.50%	40%	17.5
24h_R39_3	96.08	12513655	537909	686239	72.80%	10.5	13.9	82.20%	39%	13.7
24h_S204_1	93.82	24098191	1644396	869544	74.40%	20.6	26.7	82.60%	39%	26.6
24h_S204_2	93.73	18628563	1286915	610866	73.90%	15.8	20.6	80.50%	38%	20.5
24h_S204_3	94.24	30040681	1907176	1163330	73.60%	25.5	33.2	82.70%	39%	33.1
24h_S39_1	94.5	31169384	1871060	961437	74.20%	26.1	34.1	83.10%	38%	34
24h_S39_2	94.37	17382165	1080497	731097	74.40%	14.9	19.2	83.40%	40%	19.2
24h_S39_3	94.64	19342903	1139986	794472	75.60%	16.8	21.3	84.50%	40%	21.3
4h_R204_1	94.68	28453615	1673942	1337935	76.30%	25.2	31.5	83.70%	39%	31.5
4h_R204_2	94.69	29076780	1708128	1376727	75.80%	25.6	32.2	83.60%	39%	32.2
4h_R204_3	94.79	25150220	1445893	1177881	75.90%	22.1	27.8	83.10%	39%	27.8
4h_R39_1	95.17	21751781	1150038	902496	77.00%	19.1	23.8	81.30%	39%	23.8
4h_R39_2	95.35	22553102	1146134	964314	77.60%	20	24.7	81.80%	39%	24.7
4h_R39_3	95.26	20277893	1052779	861935	77.40%	17.9	22.2	81.00%	39%	22.2
4h_S204_1	94.51	20129434	1212317	736994	75.60%	17.3	22.1	82.40%	38%	22.1
4h_S204_2	94.75	19278863	1109330	742163	76.50%	16.8	21.2	82.60%	38%	21.1
4h_S204_3	94.73	22795207	1316462	854022	76.00%	19.7	25	83.30%	38%	25
4h_S39_1	94.59	16417993	964980	461120	77.20%	14.2	17.9	81.10%	38%	17.8
4h_S39_2	94.81	48312953	2721139	1407734	77.60%	42	52.5	91.90%	38%	52.4
4h_S39_3	94.6	18018610	1058348	529545	77.20%	15.6	19.6	81.60%	38%	19.6
90m_R204_1	94.92	25892001	1452879	1249189	76.00%	22.8	28.6	83.90%	39%	28.6
90m_R204_2	94.81	24738724	1418955	1194370	76.00%	21.8	27.4	83.30%	39%	27.4
90m_R204_3	94.85	24218779	1375789	1145706	75.70%	21.2	26.8	83.20%	38%	26.7
90m_R39_1	95.4	26243403	1314514	1009307	76.80%	22.8	28.6	85.60%	39%	28.6
90m_R39_2	95.36	28694919	1447874	1080092	76.50%	24.8	31.3	86.10%	39%	31.2
90m_R39_3	95.35	12822905	650291	510010	76.30%	11.1	14	80.90%	39%	14
90m_S204_1	96.57	14650473	542446	622413	77.70%	12.8	16	83.10%	39%	15.8
90m_S204_2	96.56	13817328	513512	586351	77.40%	12	15.1	82.40%	39%	14.9
90m_S204_3	96.54	13229150	494464	562965	77.50%	11.5	14.5	81.50%	39%	14.3
90m_S39_1	95.64	13490224	634544	438990	78.00%	11.8	14.8	80.90%	39%	14.6
90m_S39_2	95.8	14845416	672476	490901	78.90%	13.1	16.3	82.00%	39%	16
90m_S39_3	95.74	12626382	581139	441950	79.20%	11.2	13.9	81.20%	40%	13.6

3.4.3 RNA-Seq multivariate analysis

Counts for features produced by Hisat2 and HTSeq counts were utilised to produce estimates of the abundance of each gene within each of the samples. DESeq2 models the gene counts following a negative binomial distribution and applies normalisation factors to account for differences in sequencing depth between samples. These normalised gene counts were then further used to calculate the distances between our samples. We then projected the two dimensions that represented the largest distance between our samples. The dimensional reduction of multifactorial data common referred to principal component analysis (PCA) allows for the differentiation of the samples with largest difference in its variables (PC1) and second largest difference (PC2). We visualised the component spaces of the 1st three components.

From the first component it is possible to see that there is little variance between our 3 replicates per sample and that the largest separation, i.e., the x axis PC1 representing 34% of the explained variance, is due to difference in time variable of biofilm formation primarily between early (90min and 4h) and late-stage (24h) biofilm formation (Figure 3.4A). Visualising the ordination of the 1st and 3rd components reveals larger separation of the components according to the two medias over the 3rd component (Figure 3.4B). Similarly, visualising the 2nd and 3rd components there is visible distinction and clustering between the two strains across the 2nd component and 3rd component according to whether they were grown FCS or RPMI. The PC2 and PC3 account for 19.92% and 15.32% of the variability observed within the samples, respectfully (Figure 3.4C). The first 3 components are responsible for the >70% of the total variability as is shown within the scree plot (Figure 3.5A).

Our observations in ordination were further confirmed when visualising the relationship of our components and covariates. The first component is influenced by the time covariate. The media, and strain covariates have a much larger influence on the second and third components (Figure 3.5B). In summary the difference between late and early-stage biofilms accounts for most of the variation between the samples.

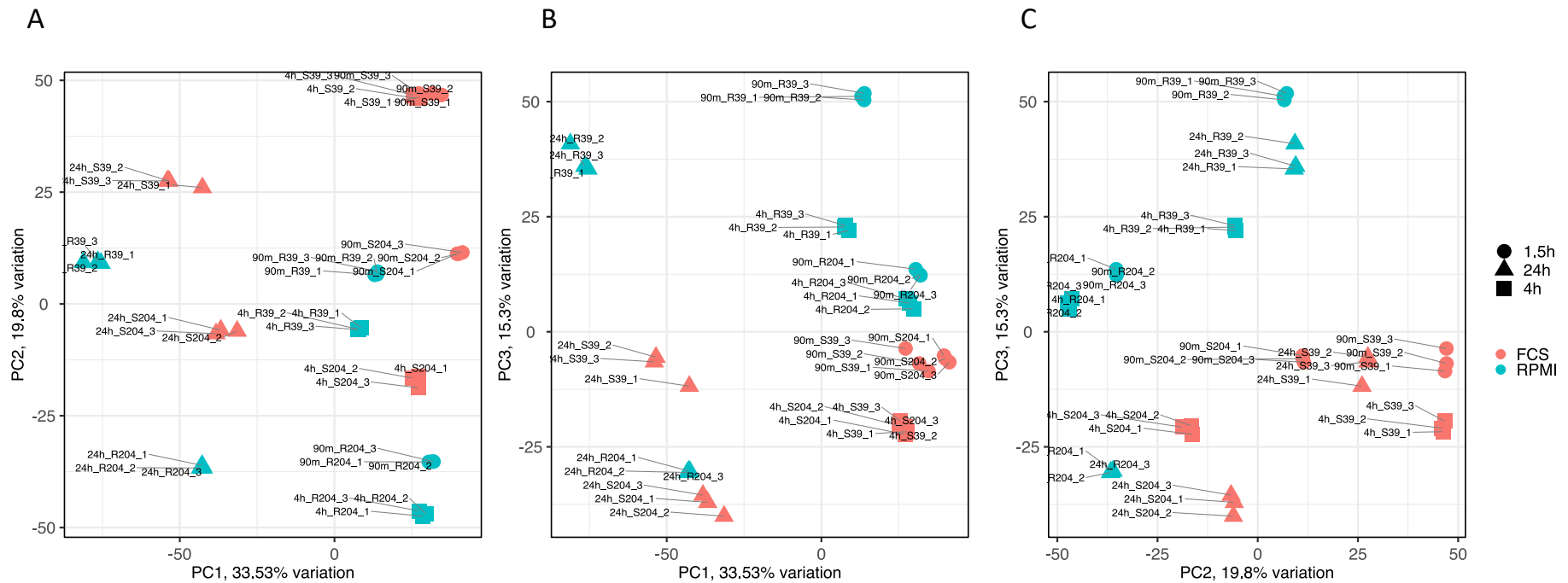


Figure 3.4 Principal Components Analysis (PCA) of *C. albicans* transcriptome. Gene count data was submitted to DESeq2 for normalisation before the data was reduced to a two-dimensional space within R. This data was then subsequently visualised, and samples were coloured according to whether they were grown in RPMI or RPMI supplemented with Serum (FCS). Additionally, a shape was overlaid on the samples according to the length of time the biofilms were grown for. This was either 1.5, 4 or 24h. Samples are also labelled in the to indicate whether they were from HBF (39) or LBF (204) strains. PCA of RNA-Seq on the HBF and LBF in the presence and absence of serum shown within a plot of PC1 /PC2 (A), PC1 /PC3 (B) and PC2/PC3 (C).

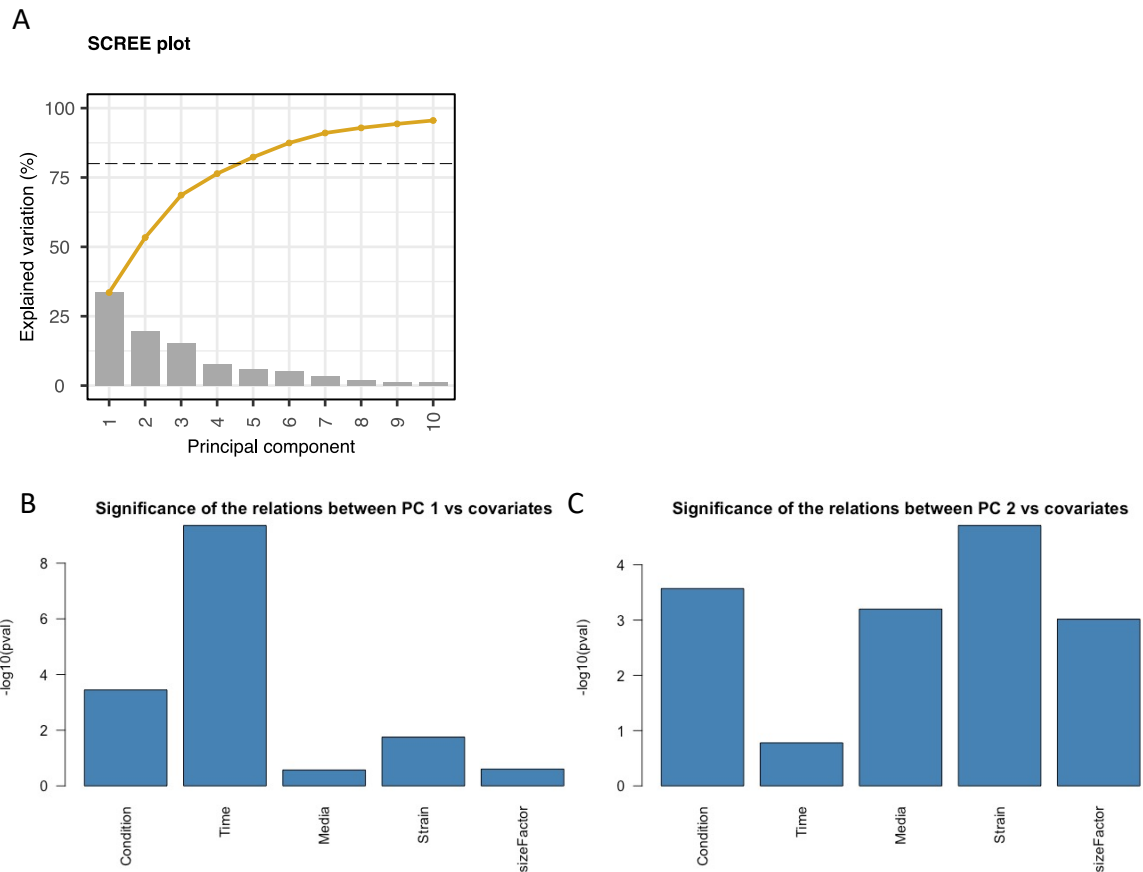


Figure 3.5 Scree plot of PCA components. Scree plot depicting components 1-10 from the principal component analysis (PCA) of all RNA-Seq data for all samples. The yellow line indicates the total explained variation, when summing components, whilst each bar indicates the percentage variation explained by that component. Dotted line indicates the point at which the total variation crosses an 80% threshold (A). Plots depicting the significance, indicated by the $-\log_{10}$ p-value, of the correlation of each grouping variable within the data versus components 1 and 2. The p-value is inverted so the higher the value the higher the significance and therefore the correlation between the covariate and that component (B).

3.4.4 *Candida albicans* differential expression in response to nutrient stress

We initially performed a pairwise comparison which only considered the grouping of samples as either grown in the presence of absence of serum. When considering the overall effect of serum on the gene expression of *C. albicans*, large numbers of differentially expressed genes were noted as discerned by using DESeq2 and grouping the samples as either serum supplemented or serum absent samples at 24h (Figure 3.6).

RPMI vs Serum

Volcano Plot RPMI = -ve and Serum(FCS) = +ve)

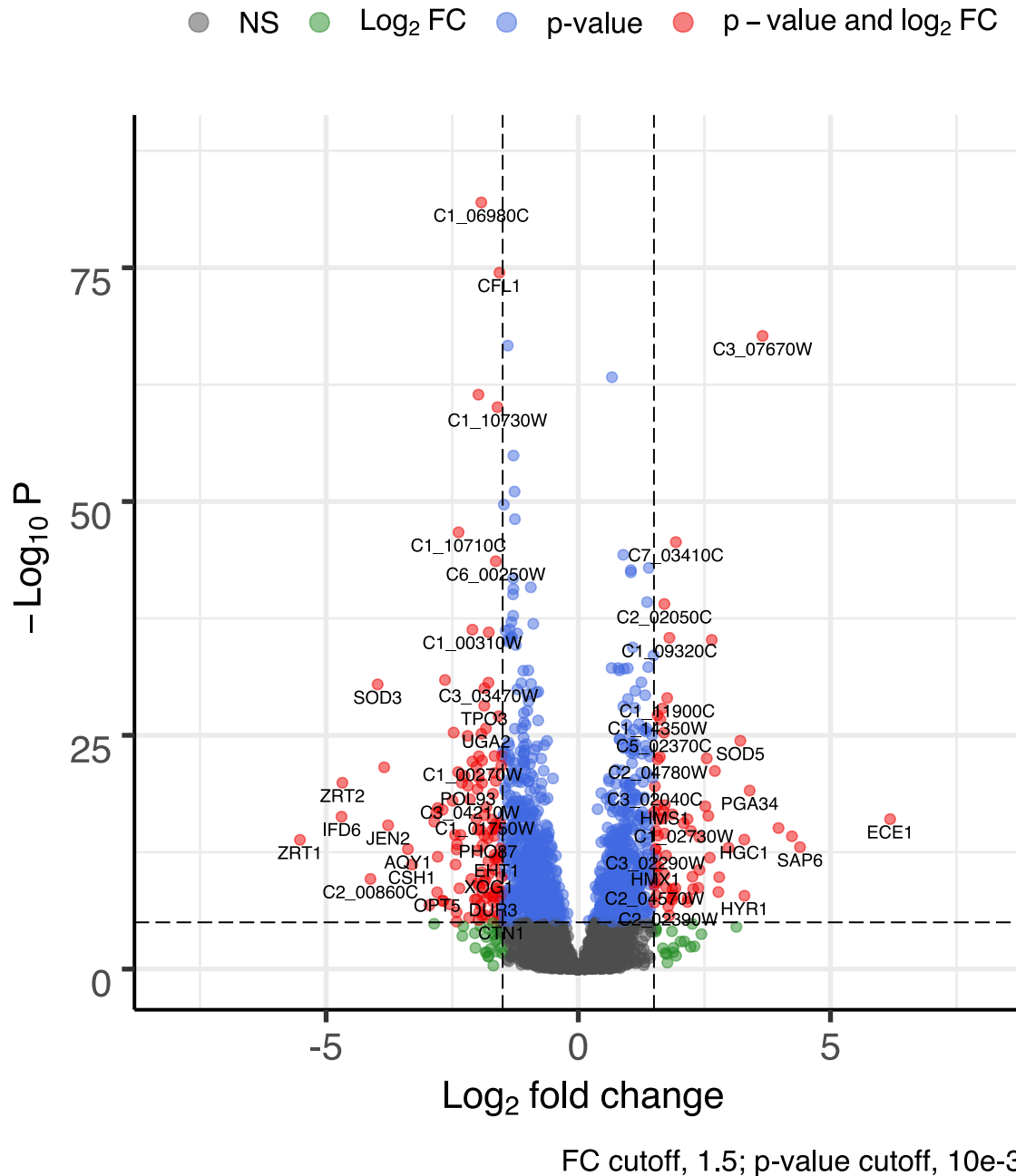


Figure 3.6 **Volcano plot depicting gross changes of *C. albicans* grown in RPMI vs Serum.** Differential expression of counts as calculated between RPMI vs Serum by DESeq2. Log₂ fold change plotted against the inverse of the log₁₀ p value. Genes upregulated in Serum are represented by a +ve Log₂ fold change and those upregulated in the RPMI only media are represented with -ve Log₂ fold change. Larger negative or positive values indicate a larger upregulation and the increased significance between the conditions is indicated by a larger negative Log₁₀ p-value (-Log₁₀P). Vertical dotted lines indicate a cut-off of Log₂FC 1.5 coloured green, horizontal dotted lines indicate a cut-off of the p-value<0.001, coloured blue or red if both cut-offs are met for illustration of the high levels of upregulated genes.

1309 genes were upregulated in strains supplemented with serum and 1184 genes were upregulated in media without supplementation, or conversely downregulated in the serum condition. Through the gene set enrichment (GSEA), which enriches based upon functional pathways, it is possible to quantify whether significant levels of differential expression across entire metabolic pathways. Gene Ontology (GO) terms and other functional groupings, such as those supplied by the KEGG database, of genes are shown significantly enriched between two conditions. We observed that between the two medias there was enrichment of numerous GO functional terms. These terms are groupings of genes belong to the same pathways and/or to the same functional groups. GSEA compared to over-enrichment analysis considers the expression profile of the all the genes of an organism and therefore considers even genes who's differential expression is modest.

GSEA identified several enriched GO pathways between our two medias. Fluctuations across multiple biological processes that include polysaccharide, lipid, glycoprotein, and protein metabolic processes were influenced (Figure 3.7A). The enrichment in pathways involved in multiple metabolic processes was mirrored when GSEA was performed against the *C. albicans* KEGG database. Fatty acid metabolic pathways which included the peroxisome, amino acid metabolism as well as carbon metabolism are all enriched within the strains grown in Serum.

These pathways maintained and curated through the KEGG database also included changes in enrichment in the sugar, protein, and lipid pathways. Serum supplementation altered the regulation of metabolic functions throughout most super pathways in *C. albicans* (Figure 3.7B). We observed that serum has a large effect on both ontological functions and the metabolic pathways in both of our strains. However, to determine which of these upregulated processes are conserved between the two strains and which are differing in HBF and LBF it was necessary to perform differential expression analysis of the strains independently in the presence and absence of serum.

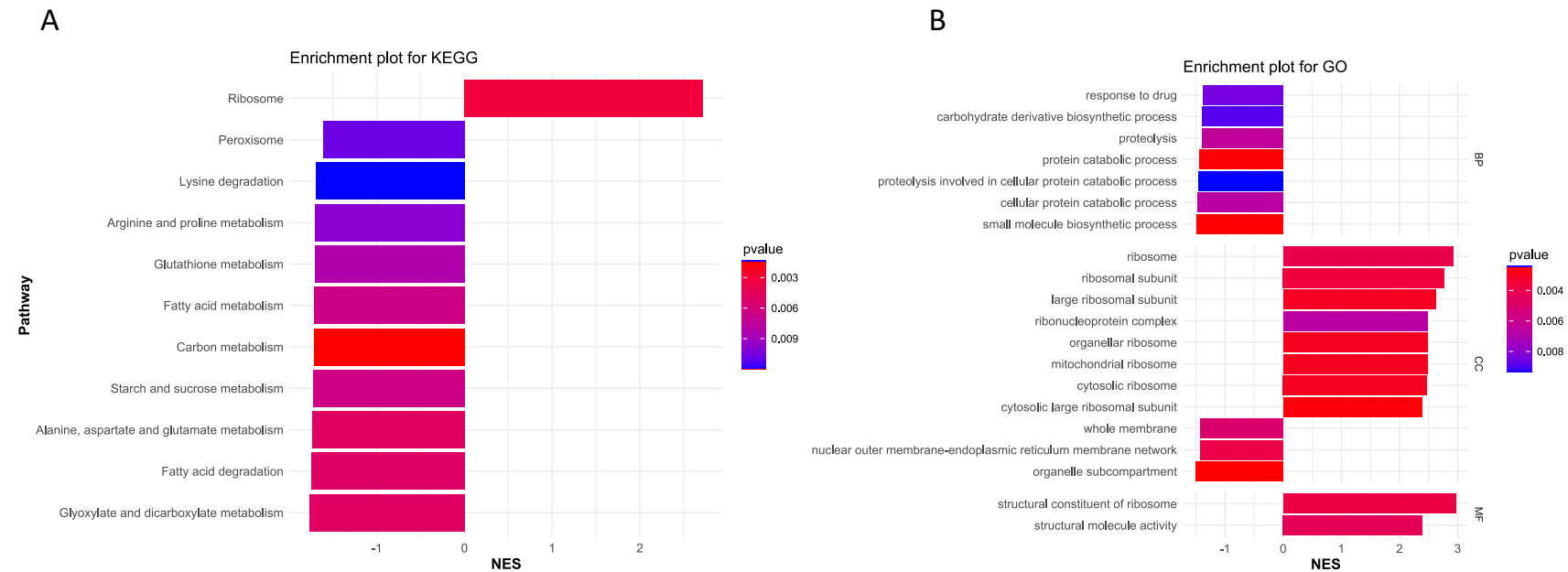


Figure 3.7 Bar plot depicting the differentially regulated Gene Ontology and KEGG pathways. Gene set enrichment analysis (GSEA) was performed on the total up and downregulated genes in the presence of serum vs the absence. Genes were ranked according to their log₂FC highest positive to lowest negative. Positive normalised enrichment score (NES) as calculated by fgsea in R indicates those pathways enriched in serum and negative enrichment indicates those upregulated in the absence of serum for both Gene Ontology pathways (A) and KEGG pathways (B). Large +ve NES indicates a greater enrichment of genes within the samples grown in serum and a smaller -ve value represents a greater enrichment of genes in the samples grown in RPMI media. The intensity of the colour indicates the significance of the enrichment according to the nominal p-value which is indicate in the legend.

3.4.5 Gene expression of strain and time dependant genes in the nutrient stress response of *Candida albicans*

Differential expression was calculated using the DESeq2 package, we performed pairwise comparisons between sample groups considering the variables media, time, and strain. Pairwise comparisons or contrasts were made between serum supplemented conditions and non-supplemented conditions for each of the strains at each of the three time points. We observed large numbers of differential expression between each of these conditions. The differences observed in HBF were much higher across each of the time conditions than the LBF, i.e., greater upregulation in RPMI and Serum were observed in HBF (Figure 3.8). Genes that had a log₂ fold change (log₂FC) ≥ 1.5 and FDR adjusted p-value of ≤ 0.05 were considered differentially expressed between our two conditions for subsequent analysis. LBF had a much lower number of differentially expressed genes, that met our cut off criteria, compared to HBF (Figure 3.8 and 3.9).

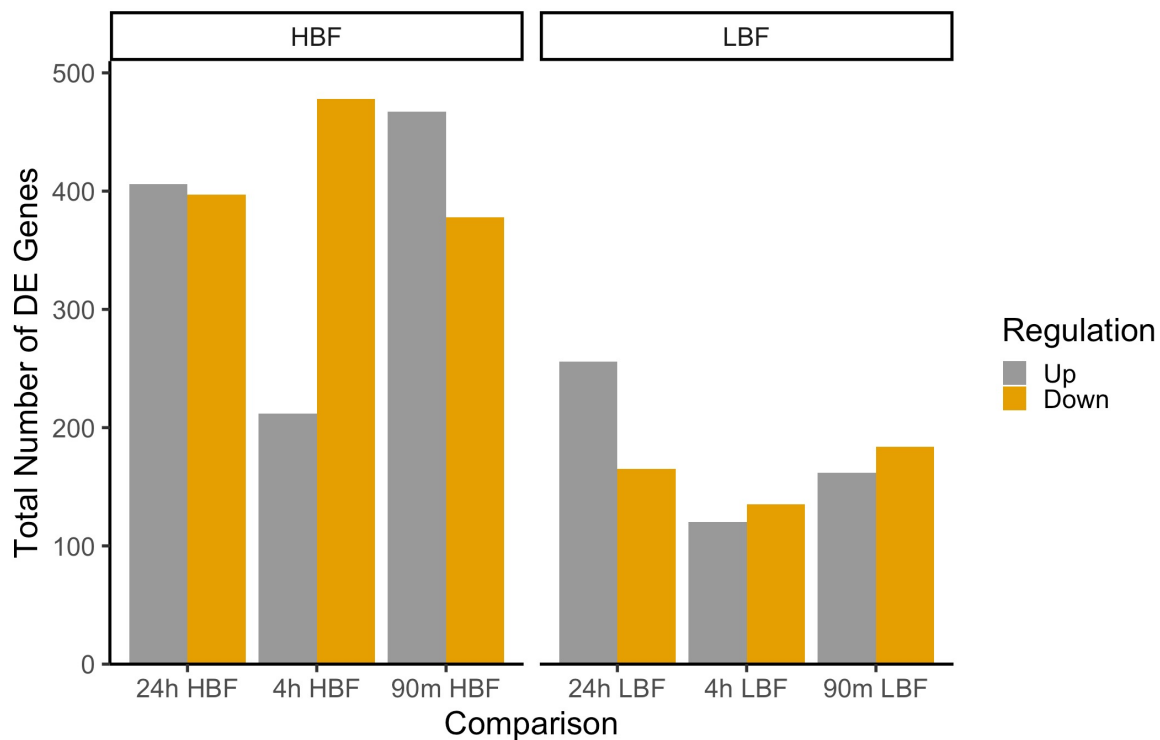


Figure 3.8 Number of up or downregulated at each time point for each strain in the presence of Serum. Up or downregulated determined based upon Benjamini-Hochberg (FDR) cut-off of ≤ 0.05 and log₂ fold change of ≥ 1.5 . Total number of genes differentially expressed are either upregulated (Up) in Serum or downregulated in Serum (Down). This comparison Serum vs RPMI was performed for the HBF and LBF strains at each time point 1.5, 4 and 24h as indicated by the x axis labels.

Chapter 3: Transcriptomic profiling of phenotypically distinct *Candida albicans* isolates

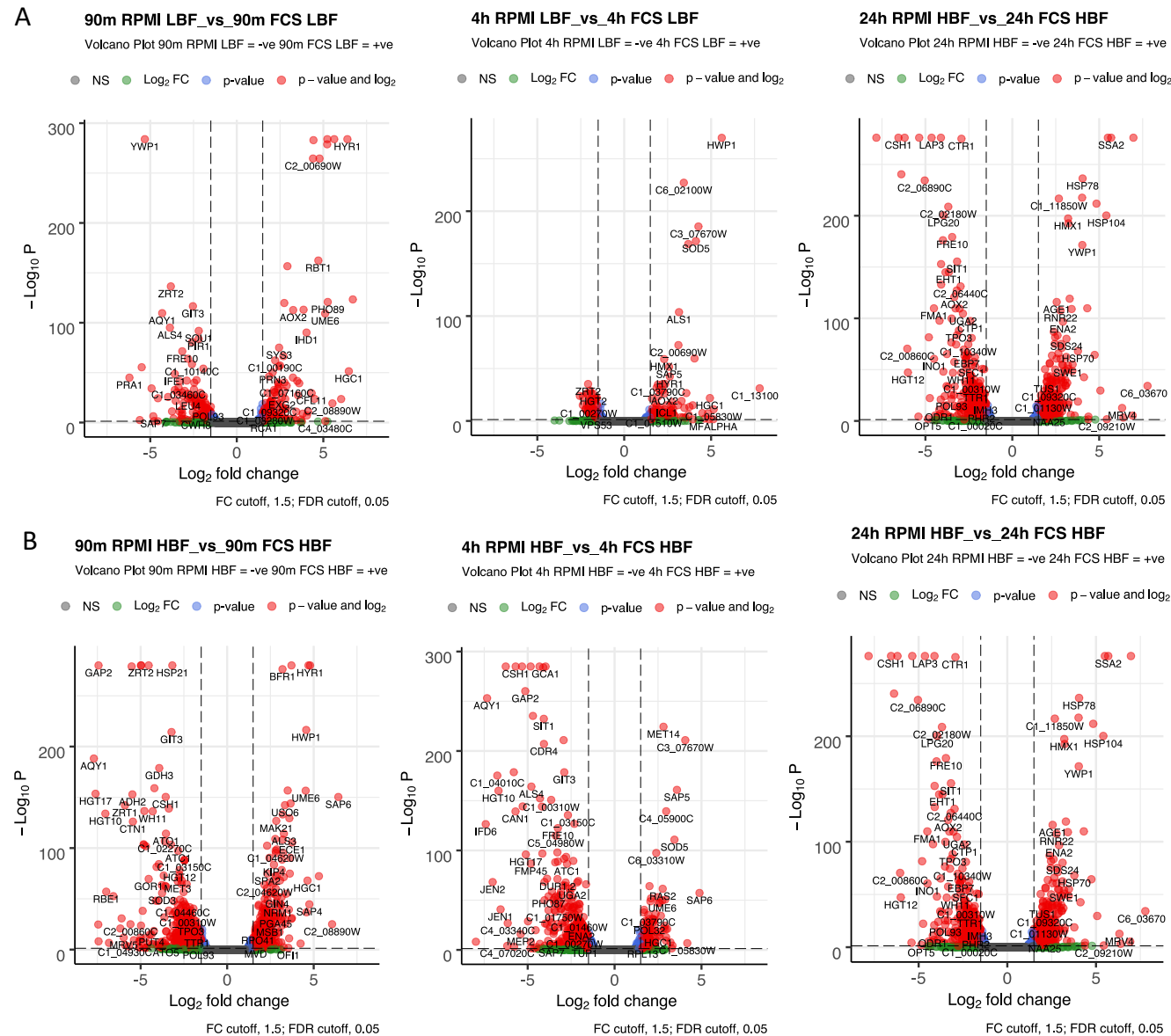


Figure 3.9 Volcano plots depicting the levels of gene expression at each time and within either the HBF or LBF *C. albicans* strain. Differential expression of counts as calculated between RPMI vs Serum by DESeq2. Log2 fold change plotted against the inverse of the log10 p value. Genes upregulated in Serum are represented by a +ve Log₂ fold change and those upregulated in the RPMI only media are represented with -ve Log₂ fold change. Larger negative or positive values indicate a larger upregulation and the increased significance between the conditions is indicated by a larger negative Log10 p-value (-Log₁₀P). Vertical dotted lines indicate a cut-off of Log2FC 1.5 coloured green, horizontal dotted lines indicate a cut-off of the FDR p value <0.05, coloured blue or red if both cut-offs are met for illustration of the high levels of upregulated genes. This analysis is plotted for all time points 90mins/1.5h, 4h and 24h. Additionally this was performed for both the LBF (204) strain and the HBF (39) strain.

The genes expressed at each time point had little overlap between each time point. Genes that were differentially expressed in either strain at either 90min, 4h and 24h seemed to be time dependent expression and distinct to that time point (Figure 3.10A, B).

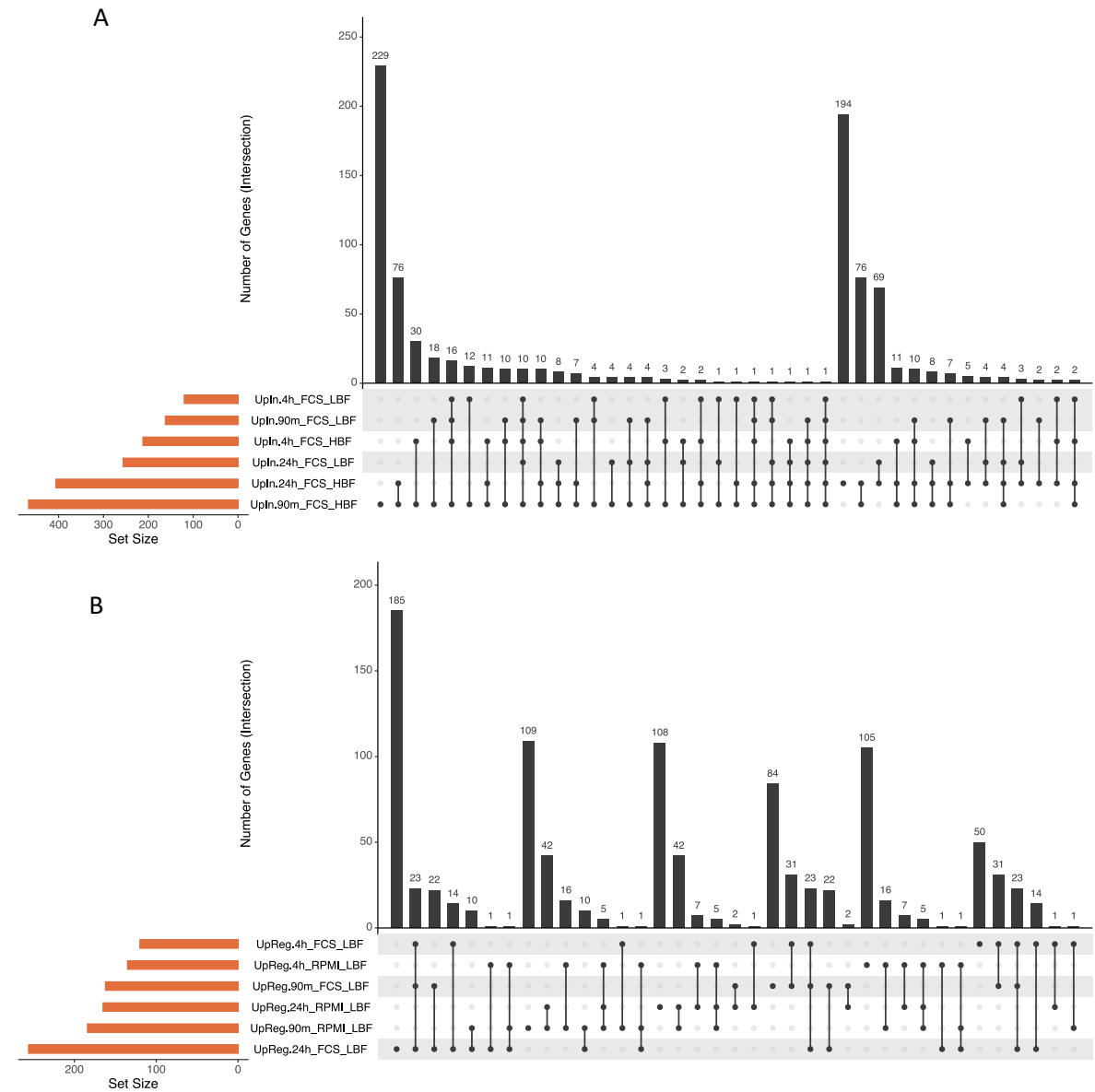


Figure 3.10 **UpSet plots depicting the overlap of differentially expressed genes.** UpSet plots performing a similar function to Venn displaying the overlap in upregulated genes between the high and low biofilm forming strains (A) or the overlap of up or down regulated genes at each time point in the LBF (B). Plot A illustrates the overlap between upregulated genes in LBF and HBF with most LBF upregulated genes being upregulated in HBF and high levels of HBF genes being unique to HBF. Plot B indicates overlap of upregulated genes in the LBF strain in serum between the timepoints with many of the genes being unique to the individual time points. Total set size for each sample, which is the total number of differentially expressed genes is also indicated in the orange bar chart. Sample names are indicated by the timepoint_media_strain.

From DESeq2 the topmost differentially up (FCS) and down (RPMI) regulated genes are shown in tables 3.3-3.4. This data mirrors the most differentially expressed genes that are annotated within the volcano plots detailed in figure 3.9. In the HBF we can see that at 90min and 4h in FCS there are several Secreted Aspartyl Proteinases (SAP) families' transcripts SAP6, 4 and 5 which are involved in early biofilm formation, virulence, and nutrient acquisition. Hyphal wall related protein transcripts such as hyphal regulated cell wall protein 1 (HYR1) and hyphal wall protein (HWP1) are also upregulated in FCS in the early biofilm time points. Top genes in the later 24h biofilm with annotation where the spore wall formation SPO75 and marvel domain protein of unknown function MRV4 and the heat shock proteins SSA2, HSP21, HSP104. Heat shock proteins are a family of proteins by yeast in response to external stimuli and stress such as heat, pH and oxidative stress (Gong et al., 2017).

The gene with the highest upregulation in the LBF grown in FCS is Endothelin-converting enzyme ECE1 at 90min and 4h time points. Following ECE1 is the biofilm related gene HWP1. Also, at these early biofilm time points are the ALS3 and SAP genes which as we have discussed in chapter 1 are typical biofilm response genes. Also, at 4h there is upregulation of the mating factor alpha MFALPHA. The MRV2 and MRV4 genes like HBF are upregulated 24h although little is known regarding their function. MFALPHA is one of the top upregulated genes also at the 24h time point.

Several common genes feature within the top downregulated genes in FCS or upregulated genes in RPMI. Including the JEN family and ZRT family genes. In the LBF within the top 10 genes the ALS2 and ALS4 are notably downregulated in FCS in contrast to the. The ZRT genes 1 and 2 are downregulated in LBF at 90 min and 24h in the top 10 genes. These genes encode the transporter for high affinity to zinc. ZRT is thought to be regulated by the extracellular levels of zinc (Zhao and Eide, 1996). These genes also feature in the top downregulated genes in FCS in the HBF strain. With the JEN2 and JEN1 genes being in the top 10 at 90min and the ZRT1 and ZRT2 appearing in the top downregulated in 4h and 24h. There is a conserved response of downregulation of carboxylic acid transport through JEN1/2 and zinc metal transport through PRA1 and ZRT1/2 in the FCS (Table 3.3-3.4).

Chapter 3: Transcriptomic profiling of phenotypically distinct *Candida albicans* isolates

Table 3.3 DESeq2 results of the top 10 up and top 10 down regulated genes in HBF. Differential expression of genes from biofilms grown for 90min, 4h or 24h in RPMI or RPMI supplemented with serum. Table includes the candida genome database identifier (GeneID), common name or gene symbol (Symbol), the FDR adjusted p value (pvalue adj) from DESeq2 and the log2 transformed fold change (Log2FC) between RPMI and RPMI+serum (FCS).

	90min				4h				24h			
	GeneID	Symbol	Pvalue (adj)	log2FC	GeneID	Symbol	Pvalue (adj)	Log2FC	GeneID	Symbol	Pvalue (adj)	Log2FC
Up in FCS	C6_02710C_A	SAP6	4.72E-151	6.423358436	C6_02710C_A	SAP6	5.46E-58	4.884388528	C6_03670C_A	C6_03670C	6.97E-35	7.780490214
	C2_08890W_A	C2_08890W	9.62E-26	6.062675904	C1_05830W_A	C1_05830W	1.73E-07	4.182630961	C7_00350C_A	C7_00350C	<2.21E-308	6.968574092
	CR_02750C_A	PGA34	4.67E-73	5.300650119	C3_07670W_A	C3_07670W	1.36E-211	4.06718883	C4_03670W_A	SPO75	6.24E-06	6.926200703
	C1_13450W_A	HYR1	<2.21E-308	4.792270742	C6_03030W_A	SAP5	1.18E-161	3.58742123	C4_02380W_A	C4_02380W	0.000530833	6.389011412
	C6_03500C_A	SAP4	2.30E-45	4.747467817	C2_00680C_A	SOD5	1.44E-111	3.444329996	C5_04210C_A	MRV4	2.15E-13	6.318641816
	C6_03030W_A	SAP5	<2.21E-308	4.698521663	C1_13100W_A	C1_13100W	4.81E-05	3.332871256	C1_04300C_A	SSA2	<2.21E-308	5.688532028
	C1_00780C_A	HGC1	1.02E-68	4.640881655	C4_03480C_A	C4_03480C	0.032170593	3.288498371	C2_04010C_A	HSP21	<2.21E-308	5.498480467
	C1_05920W_A	C1_05920W	1.74E-37	4.582357645	C6_00840W_A	GPX2	2.42E-08	3.215408513	C2_09210W_A	C2_09210W	0.022468958	5.438041397
	C4_03570W_A	HWP1	4.35E-217	4.565684695	CR_02750C_A	PGA34	1.01E-21	3.025994834	CR_08250C_A	HSP104	5.60E-201	5.410672893
	C1_06280C_A	UME6	2.96E-157	4.536268695	C4_05900C_A	C4_05900C	3.97E-140	2.970026577	CR_08270W_A	CR_08270W	2.42E-30	5.066657664
Up in RPMI	C5_03060C_A	TNA1	1.41E-58	-9.693141139	C4_06980W_A	PRA1	3.86E-185	-11.71600035	C4_06980W_A	PRA1	4.94E-196	-11.0668664
	C4_04030W_A	JEN2	7.15E-65	-8.262357034	C4_06970C_A	ZRT1	<2.21E-308	-11.00971055	C1_04010C_A	C1_04010C	<2.21E-308	-9.269299535
	C4_00430W_A	MEP2	7.24E-20	-8.119020275	C2_02590W_A	ZRT2	<2.21E-308	-9.033631435	C1_04140W_A	IFD6	<2.21E-308	-8.757298357
	CR_02920C_A	AQY1	5.68E-189	-7.704062616	C2_00860C_A	C2_00860C	3.50E-123	-8.11163529	C4_06970C_A	ZRT1	<2.21E-308	-8.65377271
	C4_01070W_A	HGT17	3.02E-154	-7.606661757	C7_03560W_A	C7_03560W	6.40E-09	-7.982027821	C6_03170C_A	MDR1	<2.21E-308	-7.823698856
	C3_06580W_A	JEN1	1.97E-25	-7.448788164	C1_04140W_A	IFD6	3.49E-127	-7.415269393	C1_04020C_A	CSH1	<2.21E-308	-6.554301007
	C3_05580C_A	GAP2	<2.21E-308	-7.434808063	CR_02920C_A	AQY1	7.72E-254	-7.338728503	C7_00110W_A	SOD3	3.90E-241	-6.386646606
	C1_05830W_A	C1_05830W	7.63E-09	-7.432085498	C4_04030W_A	JEN2	5.94E-69	-7.016540756	C2_02590W_A	ZRT2	<2.21E-308	-6.195046304
	C6_03790C_A	HGT10	1.21E-134	-7.03880208	C1_04010C_A	C1_04010C	5.24E-176	-6.75041636	C2_00860C_A	C2_00860C	3.86E-71	-6.044519982
	C1_14120C_A	RBE1	8.53E-58	-6.98895413	C6_03790C_A	HGT10	1.03E-160	-6.681095023	C7_00280W_A	HGT12	5.82E-48	-6.00751274

Chapter 3: Transcriptomic profiling of phenotypically distinct *Candida albicans* isolates

Table 3.4 DESeq2 results of the top 10 up and top 10 down regulated genes in LBF. Differential expression of genes from biofilms grown for 90min, 4h or 24h in RPMI or RPMI supplemented with serum. Table includes the candida genome database identifier (GeneID), common name or gene symbol (Symbol), the FDR adjusted p value (pvalue adj) from DESeq2 and the log2 transformed fold change (Log2FC) between RPMI and RPMI+serum (FCS).

	90min				4h				24h			
	GeneID	Symbol	Pvalue (adj)	log2FC	GeneID	Symbol	Pvalue (adj)	Log2FC	GeneID	Symbol	Pvalue (adj)	Log2FC
Up in FCS	C4_03470C_A	ECE1	<2.21E-308	12.67869901	C4_03470C_A	ECE1	3.38E-245	8.105597402	C6_03600C_A	C6_03600C	2.37E-296	9.442059298
	C4_03570W_A	HWP1	<2.21E-308	8.255221261	C1_13100W_A	C1_13100W	8.19E-32	7.794597779	C2_08890W_A	C2_08890W	1.63E-32	6.117204395
	C6_02710C_A	SAP6	4.24E-124	6.704516738	C4_03570W_A	HWP1	1.06E-270	5.612377382	C1_04050C_A	MFALPHA	0.000844085	6.084347752
	C1_00780C_A	HGC1	3.87E-52	6.472925756	C1_05830W_A	C1_05830W	7.22E-12	5.147552394	C5_04370C_A	PGA37	1.18E-71	5.876146652
	C1_13450W_A	HYR1	<2.21E-308	6.390079623	C1_04050C_A	MFALPHA	0.02461027	5.125264634	C3_06660C_A	C3_06660C	1.25E-61	5.666000006
	CR_02750C_A	PGA34	2.86E-24	6.028796729	C1_00780C_A	HGC1	1.53E-22	4.930894311	C3_07840C_A	C3_07840C	0.020310915	5.62512508
	CR_07070C_A	ALS3	<2.21E-308	5.628778798	C4_06920C_A	CSA2	0.035162412	4.881475088	C5_04980W_A	C5_04980W	3.58E-154	5.356780975
	C2_08890W_A	C2_08890W	6.33E-20	5.526122836	C4_06560W_A	PGA15	5.80E-08	4.823173405	C5_04190W_A	MRV2	<2.21E-308	5.349207219
	C4_01940W_A	PHO89	1.54E-121	5.255040098	C2_00670C_A	C2_00670C	0.006120667	4.62664836	C5_04210C_A	MRV4	4.75E-16	5.271240515
	C2_00680C_A	SOD5	<2.21E-308	5.230464941	C7_04080C_A	C7_04080C	2.87E-08	4.495752642	C6_03280W_A	C6_03280W	0.00233792	5.069912538
Up in RPMI	C4_06980W_A	PRA1	1.01E-45	-6.204797538	C3_00220W_A	HGT19	1.42E-24	-2.551670194	C1_04140W_A	IFD6	2.16E-183	-10.35092483
	CR_00930W_A	ATO10	0.003265017	-5.605518716	C6_04380W_A	ALS2	7.81E-16	-2.382991051	C4_06980W_A	PRA1	9.35E-130	-9.329122823
	C3_01540W_A	C3_01540W	3.91E-56	-5.495682951	C6_04130C_A	ALS4	6.83E-24	-2.382118599	CR_09390C_A	CR_09390C	1.60E-23	-8.776433392
	C2_08590W_A	YWP1	<2.21E-308	-5.316051485	CR_03810W_A	PRP13	9.42E-12	-2.232144112	C1_04010C_A	C1_04010C	<2.21E-308	-8.638735101
	C4_06970C_A	ZRT1	6.33E-35	-4.924355183	C7_00110W_A	SOD3	3.70E-12	-2.203643359	C4_06970C_A	ZRT1	<2.21E-308	-6.818469056
	C1_04870W_A	SAP7	4.64E-07	-4.825915327	CR_08420W_A	CR_08420W	4.55E-16	-2.146439929	C6_03170C_A	MDR1	<2.21E-308	-6.661184653
	C1_04930C_A	C1_04930C	0.029925725	-4.774792288	C3_03500W_A	SKN2	3.47E-27	-2.143715118	C3_02800W_A	ADH4	2.87E-126	-5.863051659
	C3_06580W_A	JEN1	0.046596771	-4.559240043	C2_09280C_A	C2_09280C	3.96E-26	-2.116924199	C2_02590W_A	ZRT2	<2.21E-308	-5.761689161
	C4_05730W_A	C4_05730W	4.64E-25	-4.559070311	CR_10800C_A	CR_10800C	3.15E-10	-2.09767697	C5_01380W_A	CFL5	7.70E-250	-5.744757463
	CR_02920C_A	AQY1	2.30E-110	-4.311441571	C6_04650W_A	C6_04650W	1.95E-10	-2.097637894	C2_01270W_A	CHA1	1.12E-67	-4.851203764

Using the Upset plot technique, which is analogous to Venn diagrams, and the UpsetR package it was possible to visualise all the combinations of gene expression between the up and down regulated in RPMI vs FCS (Conway et al., 2017). From this it was observed that the large proportion of genes was unique to the condition with only a small proportion of the genes sharing an overlap with the other conditions (Figure 3.10A). When comparing expression between the Low and High biofilm formers it was observed that the HBF formers had many genes expressed in FCS compared to RPMI that were not observed in the LBF strain (Figure 3.10B). However, most of the genes upregulated in the LBF strains at any of the time points was also upregulated in the HBF as shown by the overlap. These seems to indicate a unique HBF expression profile that is not present in the LBF. Lower overall up and down regulation of genes within the LBF may explain some of these differences as less genes for the LBF are represented by our cut-offs. LBF shows the most drastic change in morphology in the presence of serum but a lower level of overall differential expression.

We observed differences in gene regulation between the time points and the two strains in response to Serum stimulation. We wished to investigate this further by determining the pathways that are influenced by this phenotypic change. To achieve this, we employed functional over-representation analysis. Networks of related functionally enriched pathways were calculated and visualised using GO term functional enrichment in the Cytoscape plugin GlueGO.

We began by comparing the overrepresented pathways in the LBF at 1.5, 4 and 24h to identify both serum induced and temporal changes. A large network with acyl-coA oxidase and carboxylic acid catabolic process at the centre was found to be over-represented in LBF in serum compared to RPMI. When referring to p-value throughout this it is the FDR corrected p-value for the overrepresentation of pathways in the ClueGO overrepresentation test were pathways that were considered significantly overrepresented with FDR adjusted p-value<0.05.

Fatty acid metabolism related processes make up a large component of this network predominantly at the 24h biofilm time point (Figure 3.11 A,B). Within this central network the top upregulated processes acyl-coA oxidase activity is significantly overrepresented

($p=0.0045$, nr genes=3) and fatty acid binding (($p=0.0045$, nr genes =3) with the 3 genes POX1, POX1-3 and PXP2 resulting in the enrichment of this pathway. Similarly, over enriched was the carboxylic acid catabolic process ($p=0.0046$, nr genes =17). With many genes within this pathway being non-specifically expressed including POT1, POX1, POX1-3 and ARO10 and PXP2.

Additional networks of gene pathways were over-represented, including plasma membrane and transmembrane transport were also represented within the upregulated genes in FCS in the LBF strains at 24h. Overrepresented terms included the plasma membrane acetate transporter ($p=0.006$, nr genes=5), which was enriched by the genes ATO1, ATO10, ATO9, FRP3, FRP5. This pathway was non-specific and over enriched at all time points. The pathway monocarboxylic acid transmembrane acid transport ($p=0.026$, nr genes=6) was overrepresented in this subnetwork also by the same genes with the addition of JEN1.

Fungal cell wall terms were found to be over-represented at all time points with a couple of nodes or terms being specific to later biofilms. Cell wall and membrane component terms were also enriched at all time points, except for the anchored component of the cell membrane being unique to 24h. Fungal type cell wall ($p=0.007$, nr = 30) and hyphal wall ($p=0.04$, nr genes=14) GO terms were significantly overrepresented the cell wall subnetwork. Genes of families ALS, HWP SAP and SOD were over enriched at all time points including ALS1/3, HWP1, ECE1 and SAP4/5/6.

Many enriched terms belong to either the 24h time or are considered non-specific, with only pre-ribosome ($p=0.009$, nr genes=14) being unique to 4h. Top genes enriching this pathway include BUD21, CICC, DIM 1, ECM1 and ENP1. The terms cellular meta homeostasis ($p=0.009$, nr genes=16) and dioxygenase activity ($p=0.042$, nr genes=6) were additionally specifically over-represented at 90min as shown by the two nodes although they are part of the much larger networks (Figure 3.11B).

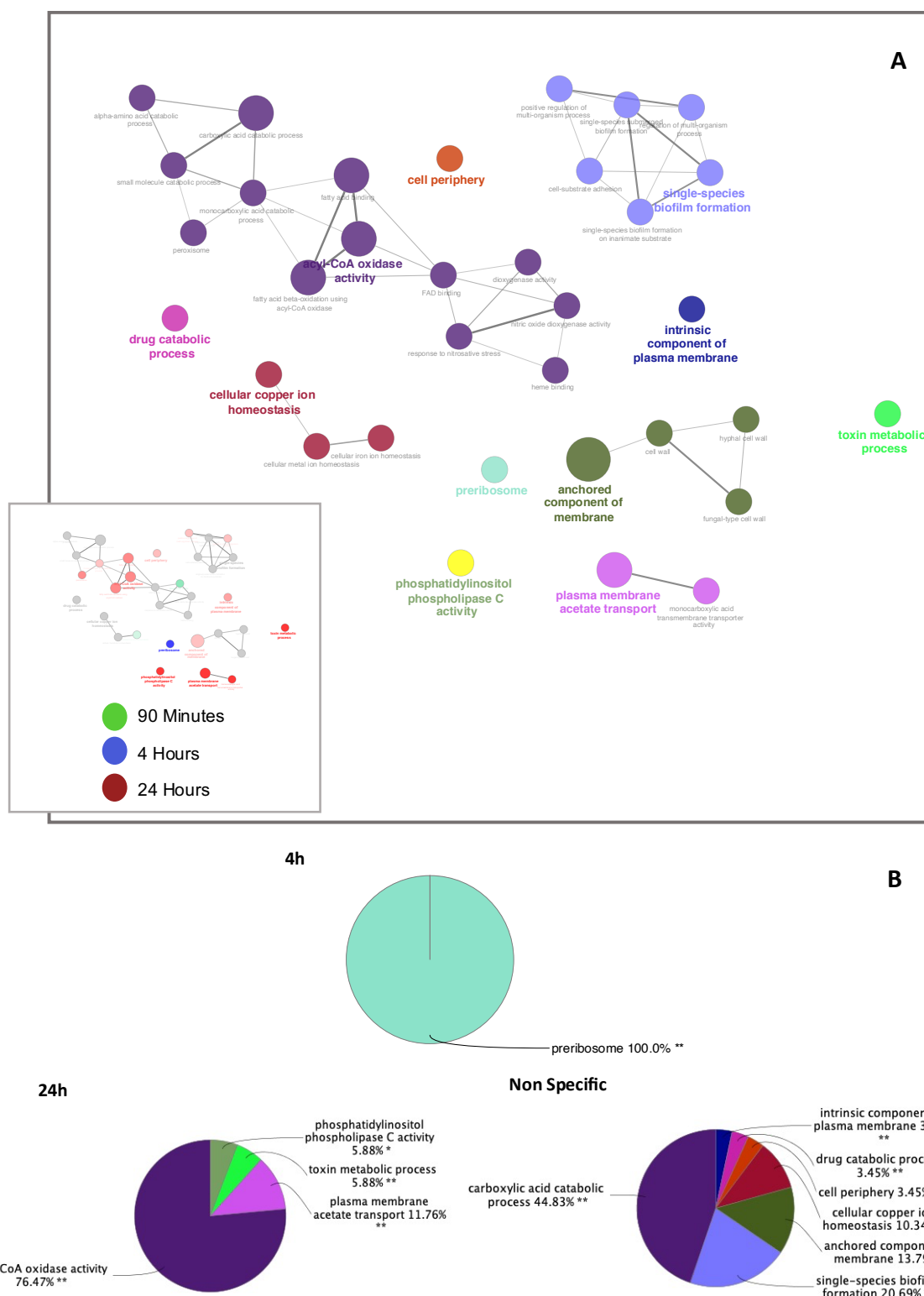


Figure 3.11 **ClueGO analysis of the upregulated genes in LBF in presence of serum and their associated pathways.** Over representation analysis of pathway analysis performed within ClueGO representing pathways as nodes linked based upon their kappa score. Nodes are coloured to show the representation of either 90min (Green), 4h (Blue) and 24h (Red). Nodes shared by the different time points strains are coloured grey. An exploded view illustrates the pathways and subnetworks which are coloured to illustrate nodes belonging to related pathways (A) Pie charts for each time points 90min, 4h and 24h and shared features summarise the percentage of terms related to a specific subnetwork of pathways also coloured to highlight the subnetwork (B).

A similar pattern is observed when considering the down-regulated pathway constructed from the RPMI enriched genes at 1.5, 4 and 24h in the low biofilm forming strains. Many of the over-enriched pathway nodes belong to the 24h time point or are considered non-specific to a particular time point, as shown by the points coloured in red or grey. This indicates that these pathways are either activated throughout biofilm formation or specifically at the 24h time point. A small number of nodes and sub-networks of 90mins downregulated genes can be seen related to protein translocation and post replication repair. Specifically, the terms at downregulated in serum at 90min are organic ion transmembrane transporter activity and chaperone cofactor-dependent protein refolding and many of the nodes are part of the larger sub-networks comprised of non-specific and 24h genes. (Figure 3.12B).

Genes represented at 24h and non-specifically comprise nodes within the 3 largest subnetworks. A subnetwork comprised mostly of terms upregulated at 24h that includes small molecule catabolic processes, e.g., terms such as serine catabolism and other amino acid related metabolic processes ($p=0.021$, nr genes=21). Genes involved in serine metabolism downregulated include ADE6, ARO9, CHA1 and MET, SHM family of genes.

The largest sub-network and group of terms is summarised as transmembrane and iron transport and is comprised primarily of genes enriched either non-specifically, i.e., at all-time points or the 24h time point. Figure 3.12A and 3.12B. This network and linked networks show downregulation in the pathways involved in transmembrane transport, particularly with regards to ion homeostasis in LBF strains grown in FCS. Pathways involved in ion homeostasis represent the most some of the top overrepresented. The pathways ion transport ($p<0.00001$, nr genes=64), anioin transport ($p=0.0034$, nr genes=22) and ion transmembrane transport ($p<0.00001$, nr genes=48) are all overrepresented. It can be observed large numbers of genes related to ion transport are downregulated. Common genes include AGP2 and ATO10. ZRT2 is also a common gene within these pathways and is one of top downregulated in Serum. These and other genes are related to Zinc transport and other metal ion transport.

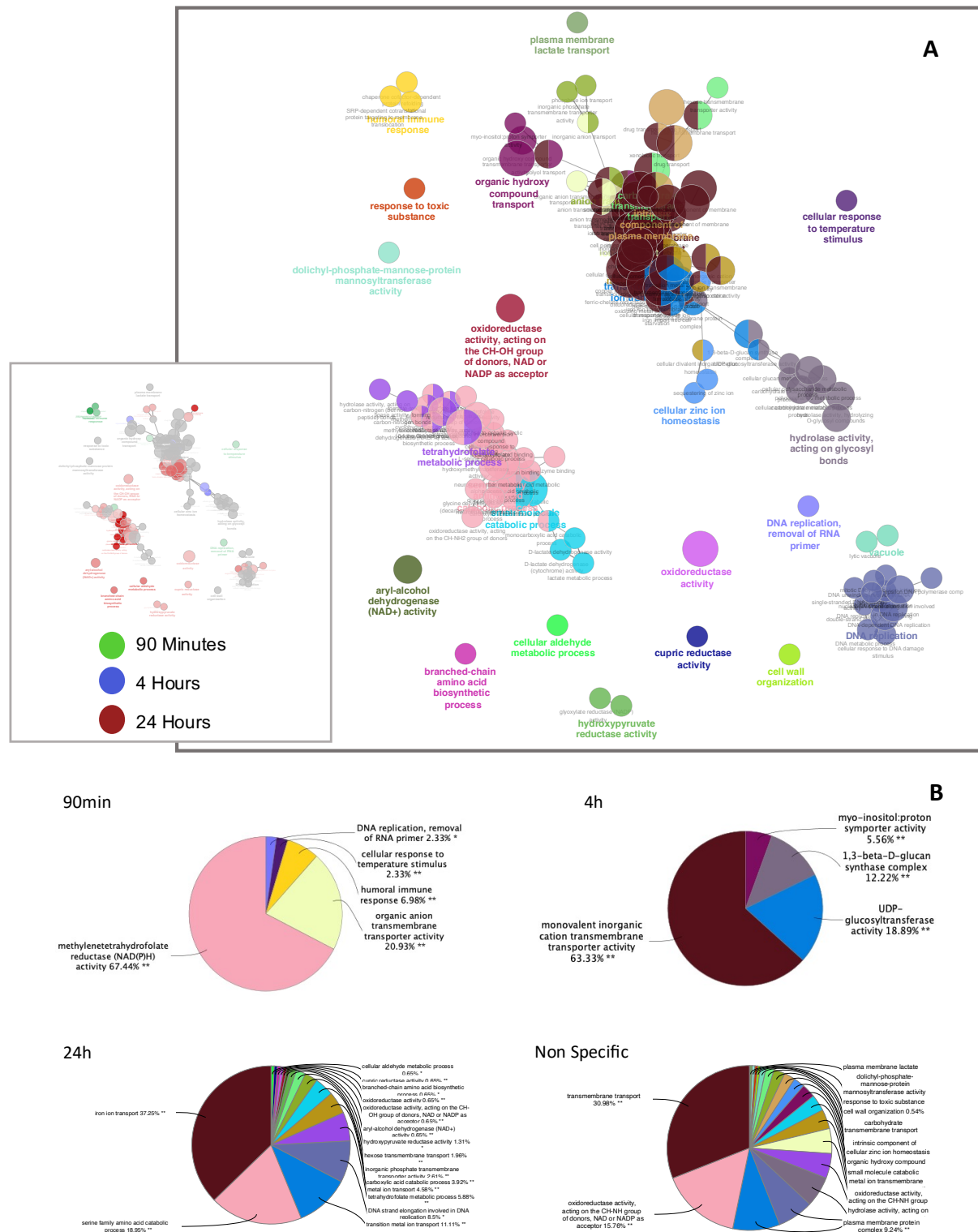


Figure 3.12 ClueGO analysis of the downregulated genes in LBF in presence of serum and their associated pathways. Over representation analysis of pathway analysis performed within ClueGO representing pathways as nodes linked based upon their kappa score. Nodes are coloured to show the representation of either 90min (Green), 4h (Blue) and 24h (Red). Nodes shared by the different time points strains are coloured grey. An exploded view illustrates the pathways and subnetworks which are coloured to illustrate nodes belonging to related pathways (A). Pie charts for each time points 90min, 4h and 24h and shared features summarise the percentage of terms related to a specific subnetwork of pathways also coloured to highlight the subnetwork (B).

Many terms related to transmembrane transport are also downregulated in the LBF strains in Serum. These include integral component of the membrane ($p < 0.00001$, nr genes=152) and the transmembrane transport ($p < 0.00001$, nr genes=84). These terms are overrepresented at all time points. Transmembrane transport and ion transport are represented by brown nodes and comprise a large proportion of the overrepresented terms in the network. Pink nodes. The pink node related to serine and other amino acid metabolism is also comprised of oxireductase terms.

When comparing the enriched terms in the high biofilm forming strain in FCS compared to the LBF there are numerous more nodes enriched in the HBF (Figure 3.13-3.15). In total there are 80 significantly overrepresented pathways FDR p-value < 0.05 with 79 of these belonging to the HBF. Only nitric oxide dioxygenase ($p = 0.0066$, nr genes=3) activity was specific to LBF and there was no overlap in non-specific nodes between the LBF and HBF at 90min. There are large subnetworks of genes overrepresented in HBF that are not present in the LBF strain. Covering many areas of cell regulation including transcriptional control, nucleolus processes, microtubule and cytoskeleton and binding. Ras transcriptional control was overrepresented in HBF as well as single species biofilm. Indicating biofilm specific responses that are present at 90 mins that are not present in the LBF or at least at high levels. The top overrepresented pathways include genes involved in nucleolus ($p < 0.00001$, nr genes=49), ribosome biogenesis ($p < 0.00001$, nr genes=49) and rRNA processing ($p = 0.000042$, nr genes=38).

Additionally at 90 min the subnetworks were comprised for the genes upregulated in FCS for the HBF strain including cell adhesion and biofilm formation related terms, regulation of transcription terms, nucleolus, pre-ribosome and protein modification (Figure 3.13-3.15).

At 4h, shown in Figure 3.14 there is more representation of nodes from the LBF strains compared to the 90min time point shown in Figure 3.13. There are several subnetworks containing terms related to ribosome and pre-ribosome, nitric acid dioxygenase activity and snRNA pseudo uridine synthesis overrepresented in the LBF. Although much larger subnetworks of overrepresented terms related to HBF similar to the 90 min network. There is a total of 37 significant pathways FDR p-value < 0.05 . 8 of these are specific to the LBF strain at 4h while 7 are non-specific, or overrepresented in both the LBF and HBF strains. The

remainder are specific to HBF strain. The top pathways in the LBF only are ribonucleoprotein complex biogenesis ($p=0.0001$, nr genes =30 genes) and the preribosome ($p=0.001$, nr genes =14). The top non-specific overrepresented terms include ribosomal large subunit biogenesis ($p=0.001$, nr genes =10), response to nitrosative stress ($p=0.0061$, nr genes =3). Nitrosative stress and nitric oxide catabolic process are enriched by the YHB1 and YHB5 genes that are involved in nitrosative and oxidative stress in yeasts. Additionally, of interest there are hyphal cell wall terms that are overrepresented in HBF in the presence of serum that aren't present in the LBF. These include the fungal ($p=0.009$, nr genes=19) and hyphal cell wall ($p=0.0091$, nr genes=11). However not overrepresented many of the usual hyphal and biofilm related genes including ALS1, ECE1, HWP1, and SOD5 upregulated in both HBF and LBF.

At 24h the representation of nodes and subnetworks comprised of HBF enriched terms is much more apparent than those by the LBF enriched terms shown in Figure 3.15. There are a total of 37 significantly overrepresented pathways with an FDR p value <0.05 . 5 of these pathways were specific to LBF, 1 was non-specific and 31 were specific to HBF.

LBF is terms comprise 2 singular nodes and a small subnetwork. The subnetwork is related to acyl-CoA and fatty acid oxidation activity and the two independent nodes are related to the anchored component of the cell wall and plasma membrane acetate support. The top pathways are fatty acid binding ($p=0.00075$, nr genes=3), acyl-CoA oxidase activity ($p=0.00075$, nr genes=3) and fatty acid beta-oxidation using acyl-CoA oxidase ($p=0.00075$, nr genes=3). Which all share the same 3 genes POX1, POX1-3 and PXP2. We previously observed these genes to be involved in a Serum specific response, they now seem to be a LBF specific response. Plasma membrane transport ($p=0.00078$, nr genes=5) is also overrepresented in only the LBF which is comprised of the ATO1, ATO10, ATO9, FRP3 and FRP5 genes. The functional term positive regulation of multi-organism process ($p=0.047$, nr genes= 16) was non-specific. This pathway includes genes that may be involved in influencing organisms of the same species. These genes may be involved in communication between *Candida* within the biofilm. Genes include the virulence ACE2, adhesin ALS1 and the mating pheromone alpha factor MFALPHA. A large, overrepresented sub-network and smaller subnetworks in the HBF strain in FCS at 24h is related to transcriptional regulation and control. Whilst other individual nodes and subnetworks represent mitotic cell cycle control (Figure 3.15). The top

overrepresented pathways that are specific to HBF at 24h in Serum are intracellular organelle ($p < 0.00001$, nr genes=197) and intracellular membrane organelle ($p < 0.000001$, nr genes =178). A subnetwork of transcriptional controlled processes including regulation of cellular processes ($p < 0.000001$, nr genes =125).

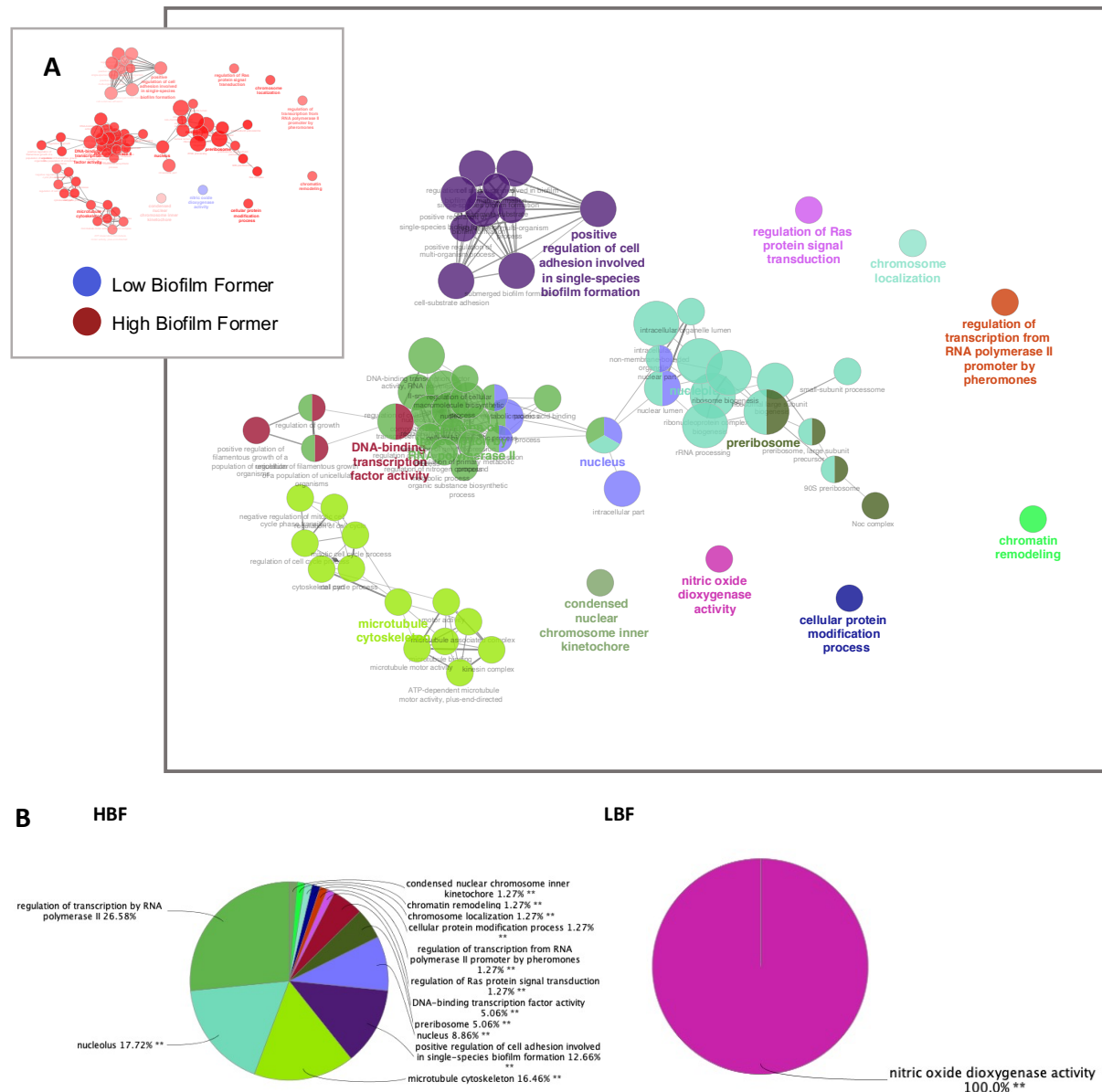


Figure 3.13 **ClueGO analysis of the upregulated genes in LBF and HBF strains in the presence of Serum at 90min.** Over representation analysis of pathway analysis performed within ClueGO representing pathways as nodes linked based upon their kappa score. Nodes are coloured to show the representation of either High (Red) and Low (Blue). Nodes shared by the two strains are coloured grey. An exploded view illustrates the pathways and subnetworks which are coloured to illustrate nodes belonging to related pathways (A). Pie charts for each strain and shared features summarise the percentage of terms related to a specific subnetwork of pathways also coloured to highlight the subnetwork (B).

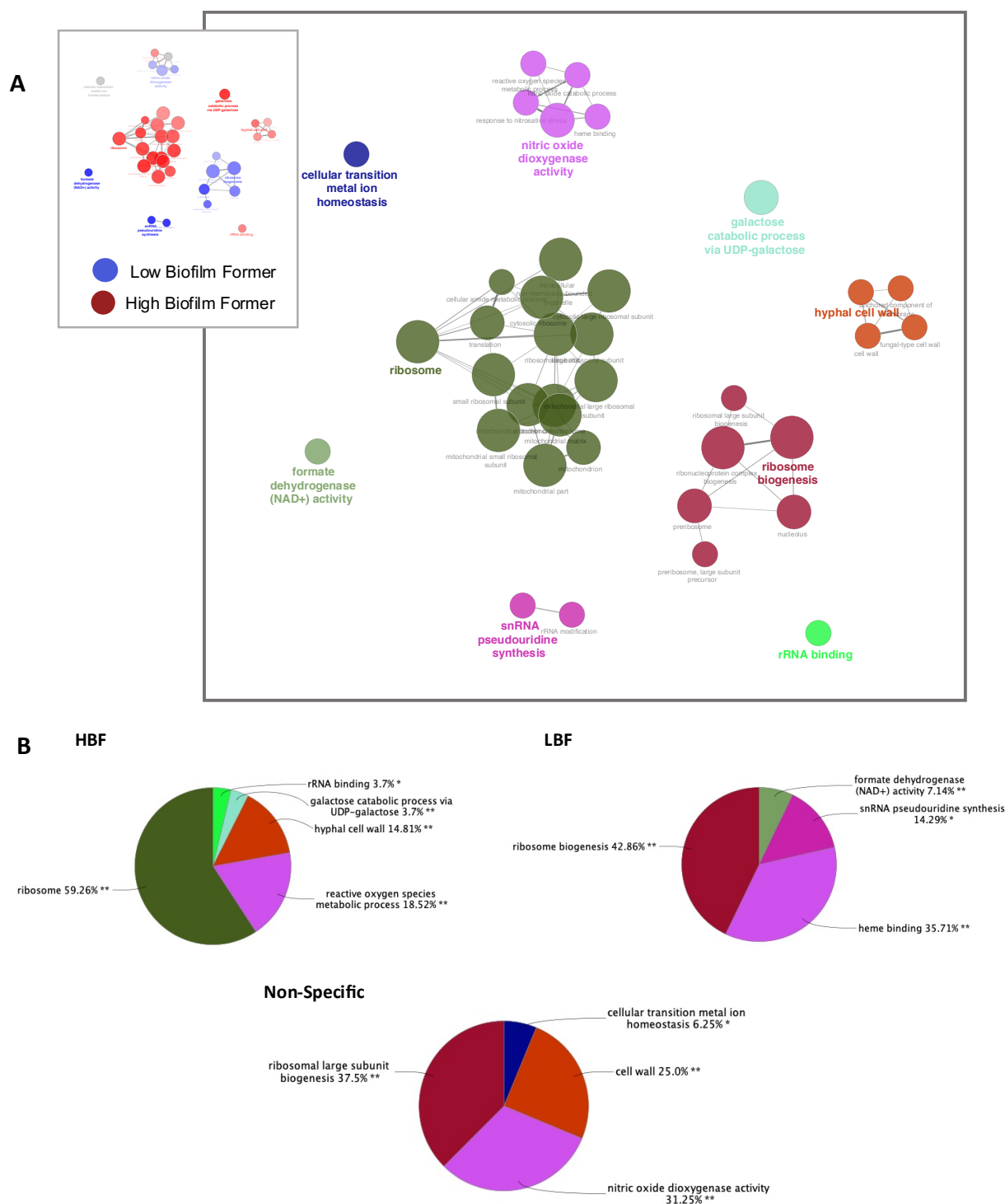


Figure 3.14 **ClueGO analysis of the upregulated genes in LBF and HBF strains in the presence of Serum at 4h.** Over representation analysis of pathway analysis performed within ClueGO representing pathways as nodes linked based upon their kappa score. Nodes are coloured to show the representation of either High (Red) and Low (Blue). Nodes shared by the two strains are coloured grey. An exploded view illustrates the pathways and subnetworks which are coloured to illustrate nodes belonging to related pathways (A). Pie charts for each strain and shared features summarise the percentage of terms related to a specific subnetwork of pathways also coloured to highlight the subnetwork (B).

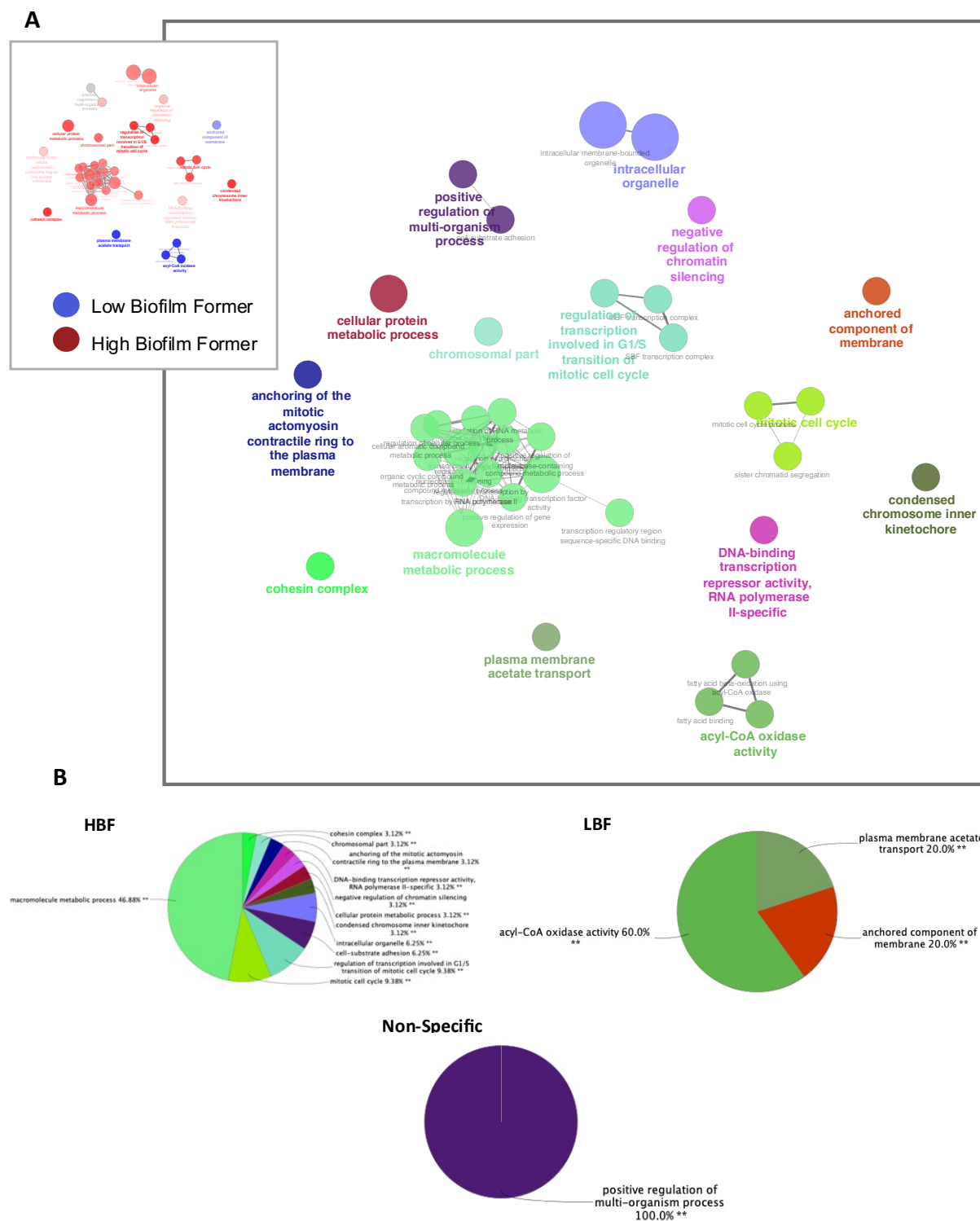


Figure 3.15 **ClueGO analysis of the upregulated genes in LBF and HBF strains in the presence of Serum at 4h.** Over representation analysis of pathway analysis performed within ClueGO representing pathways as nodes linked based upon their kappa score. Nodes are coloured to show the representation of either High (Red) and Low (Blue). Nodes shared by the two strains are coloured grey. An exploded view illustrates the pathways and subnetworks which are coloured to illustrate nodes belonging to related pathways (A). Pie charts for each strain and shared features summarise the percentage of terms related to a specific subnetwork of pathways also coloured to highlight the subnetwork (B).

3.4.6 Gene set enrichment analysis of strains in the presence of serum

GSEA utilising the entire gene set of differentially expressed genes of Serum vs RPMI at 24h were utilised to identify the top enriched pathways according to their adjusted p-values. Comparable to over representation analysis such as those performed by ClueGO the entire gene set is utilised and the genes are ranked. In this case genes were ranked according to their log2FC determined by differential expression analysis by DESeq2. The high biofilm formers display larger levels of both activated and suppressed pathways in the presence of the serum (Figure 3.16A). Larger numbers of pathways activated in HBF are unique to this strain like the ClueGO over representation analysis. However, there is a unique set of differentially enriched pathways in the LBF. 25 pathways that are enriched in the LBF in the presence of serum are unique to serum. Many of the pathways are involved in cell-to-cell fusion that and conjugation in the LBF strain. Also, the metabolic pathways of fatty acid oxidation and beta-oxidation are enriched in the LBF compared to the HBF.

When comparing the top enriched gene ontology pathways in HBF and LBF we see that most of the top pathways follow the same enrichment pattern in both LBF and HBF. The enrichment in HBF is usually higher than the LBF strains. There are several terms related to DNA replication the ribosome and nucleolus that are downregulated in LBF in serum but upregulated in HBF in serum at 24h. This concurs with our pathway analysis by ClueGo which identified significant networks of similar terms. (Figure 3.16B). Enrichment between the two strains does diverge in a few instances but many of the RPMI enriched terms are similarly enriched in both HBF and LBF. The term Oxidoreductase is downregulated in serum for the LBF strain, and it is downregulated and enriched in the absence of serum in the high biofilm formers. Transport is similarly enriched in both strains but more enriched in HBF. Metal ion terms are enriched in RPMI or downregulated in Serum in both which agrees with our previous pathway and gene expression. In this instance however the enrichment score is greater for the LBF compared to HBF (Figure 3.16B).

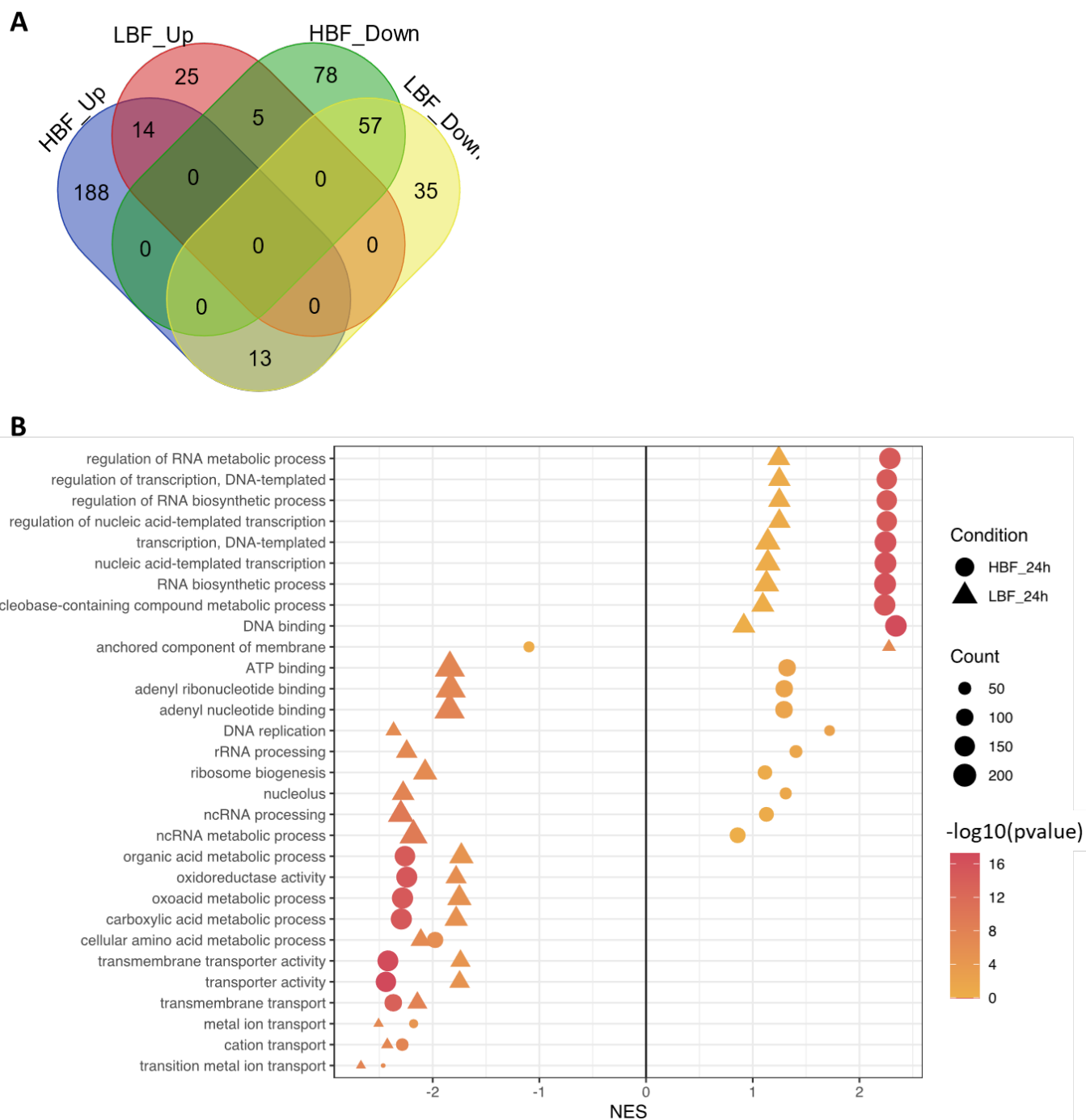
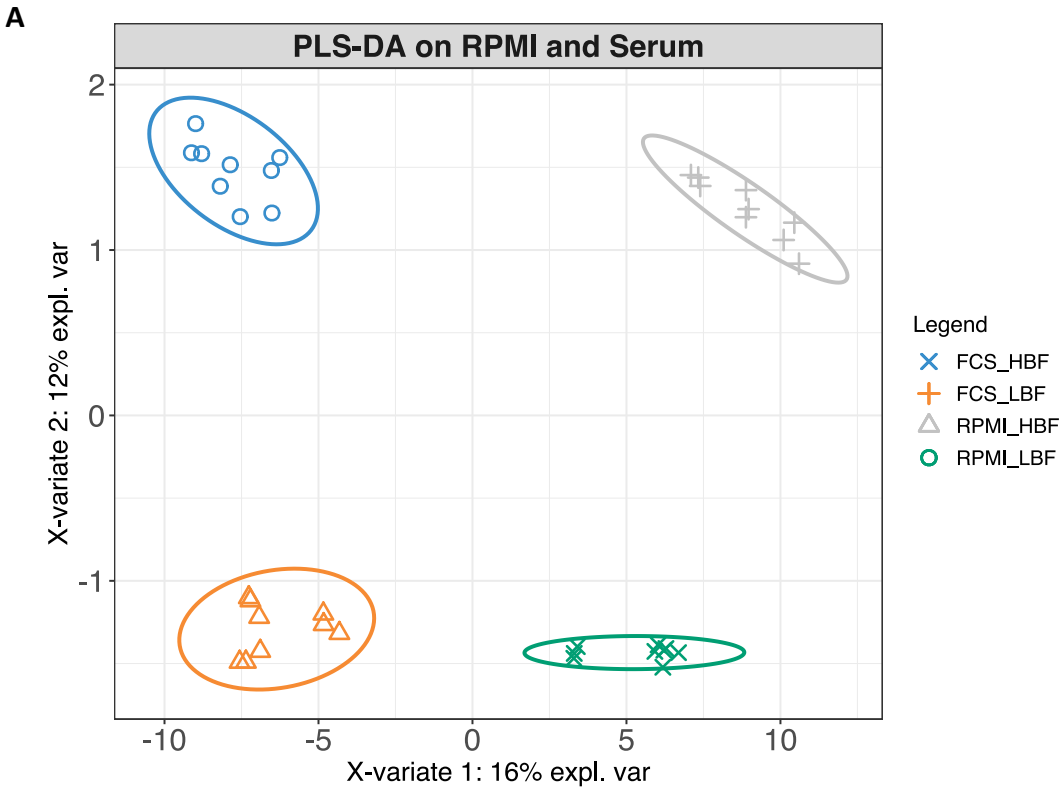


Figure 3.16 **Gene set enrichment analysis demonstrating differentially regulated pathways in the presence/absence of serum in high and low biofilm formers at 24h.** Venn diagram depicting all the enriched pathways with an adjusted p-value <0.05 in either the low or high biofilm formers (A). Enriched pathways are either up in serum or down in serum. Gene set enrichment analysis of the top enriched 15 pathways at 24h from LBF and HBF according to the Benjamini-Hochberg adjusted p-value. GSEA shows enrichment of genes in presence of serum(+NES) and downregulated genes, i.e., in enriched the absence of serum (-NES) (B). Normalised enrichment score was calculated for each pathway by the fgsea algorithm within the ClusterProfiler R package.

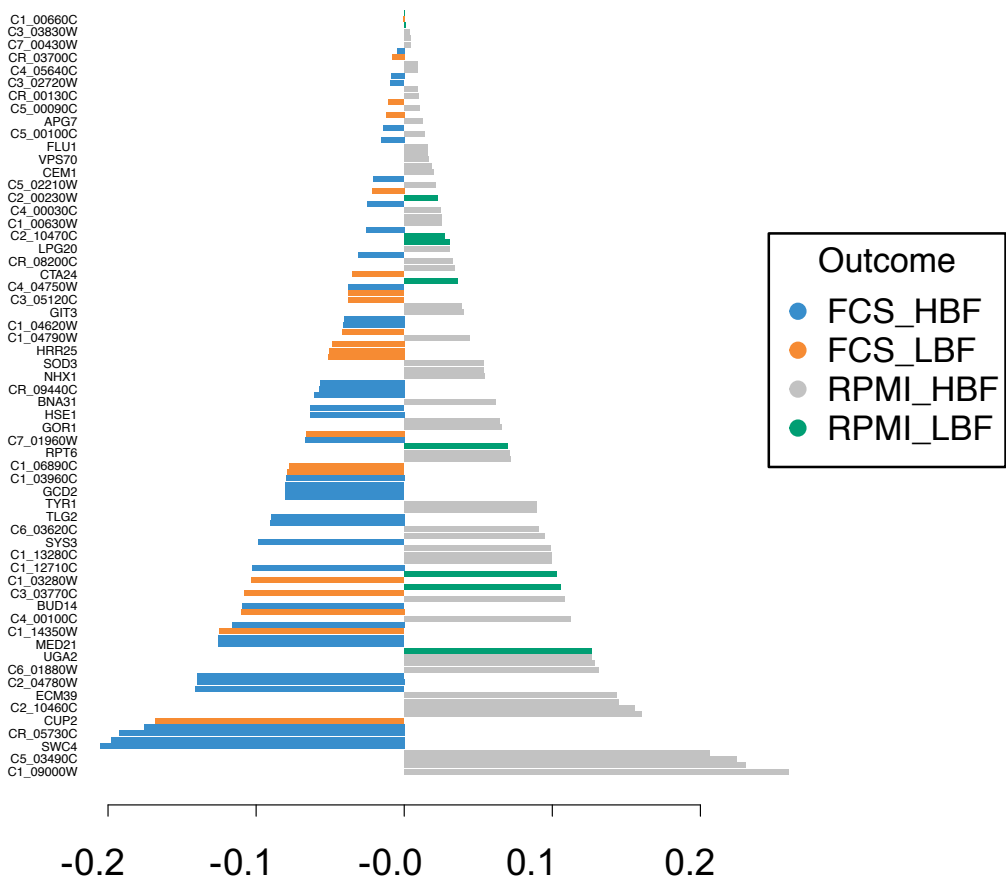
3.4.7 Biomarker analysis

Partial least squares-discriminant analysis (PLS-DA) is a method of multivariate reduction that differs from PCA in that it identifies the variables, in this case genes, that explain the highest variation in the response variable. PLS uses the response variable i.e., strain and media to maximise the variance in the different classes. Usefully in RNA-Seq it removes irrelevant or misleading variation and allows for a better prediction of markers, i.e., expressed genes that vary the most between our classes. Within the context of our analysis the PLS-DA methodology from the R package mixOmics was able to select the optimal number of genes and the most optimal number of components to depict the variation between HBF/LBF in Serum compared to RPMI. PLS-DA shows separation of the strains grown in RPMI and Serum with large separation based upon the 120 variables selected (Figure 3.17A). The relative effect on the variation on component 1 are shown with the relative effect on the variation between the strains grown in RPMI vs Serum. The loadings plot of the top 50 variables from the 1st component are shown (Figure 3.17B). It is also possible to see the distribution of our variables on component 1. Variables are those that have the most contribution to that component and the value is the loading of the sample with the highest contribution on that the component. Variables are coloured according to a group have the highest contribution or expression within that group. Indicating that they are more highly expressed within that group. The PLS-DA here classifies our samples according to the molecular signatures which are the most optimum for our identifying samples in the groups (media and time). A heatmap of our signature variables and their relative expression are shown in Figure 3.17A. Many of the genes identified are unclassified and are only indicated by their CGD accession IDs. The top discriminatory annotated genes with higher relative expression in FCS were SWC4, CUP2 and BUD14. SWC4 is of the histone acetyltransferase complex involved chromatin remodelling, CUP2 is a copper binding response and BUD14 which has a predicted role in bud site selection. Two molecular signatures were determined to explain the classification of samples on our second component between our two strains. This included a gene expressed in RPMI C7_03580C and the gene C1_12890. Both genes are unclassified within the current literature.



B

Loadings on comp 1



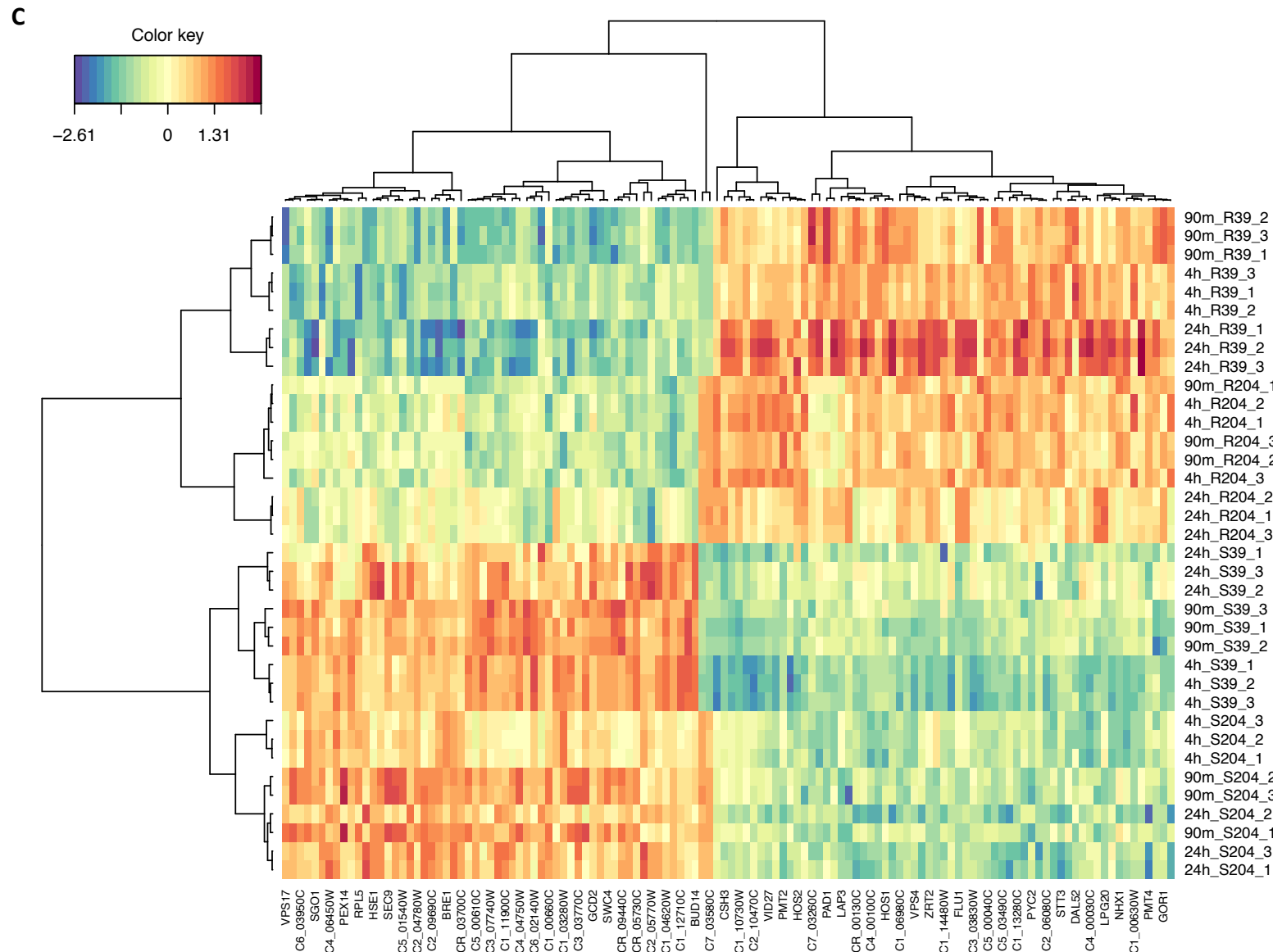


Figure 3.17 Supervised analysis of HBF and LBF strains in the presence/absence of serum using PLS-DA. Partial least squares discriminant analysis (PLS-DA) multivariate analysis displays separation of our groups based upon biomarkers showing large separation between groups (A). The mixOmics R package allows for the tuning of data to select the most discriminant variables between our groups of interest. These are the two biofilm forming phenotypes HBF and LBF grown in either RPMI or RPMI+FCS. The separation of variables according to FCS or RPMI can be seen on the x-variate. The loadings that have the largest impact are shown in the loadings plot (B) Loadings plot of the top 50 features that separate our groups on the first component and are coloured according to the sample with the largest loading for that variable. Clustered image heatmap of these discriminant features shows patterns of up and downregulated genes in RPMI and Serum (C).

3.4.8 Cell Wall Proteomics

A cell wall proteomic analysis was performed on 3 separate clinical isolates previously classified as high biofilm forming and 3 strains classified as low biofilm forming. The isolates used are summarised in section 3.3.3 in Table 3.1.

Cell wall proteomics was performed only once for every strain in each growth conditions. These were biofilms grown for 24h with or without the presence of serum. Proteins with peptides per protein ≥ 2 were quantified according to their relative peak area. Proteins were filtered so that there was consensus of at least two of our 3 strains per group with that protein present. Within the cell wall proteomics analysis, a limited number of samples were used without any technical replication. Due to this limitation proteins were considered only in regard to their presence or absence within the two phenotypic groups in the presence and absence of serum. This data was used to create Venn diagram like plots of overlaps between the different phenotypic groups (Figure 3.18A). Additionally, we visualised the log₁₀ peak area of the proteins with each sample in the form of a heatmap to summarise the presence and absence levels between our two phenotypes and our two growth conditions (Figure 3.18). We hypothesised that there may be differences in the cell wall composition between high and low biofilm formers in how they remodel their cell walls in the presence of serum. However, due to limitations in the procedure, for example, its ability to discern only the cell wall proteins and the lack of technical replicates it was deemed inappropriate to form any statistical inference. For this reason, visual analysis only was performed, and the experiment was used for hypothesis formation only with no strong conclusions possible.

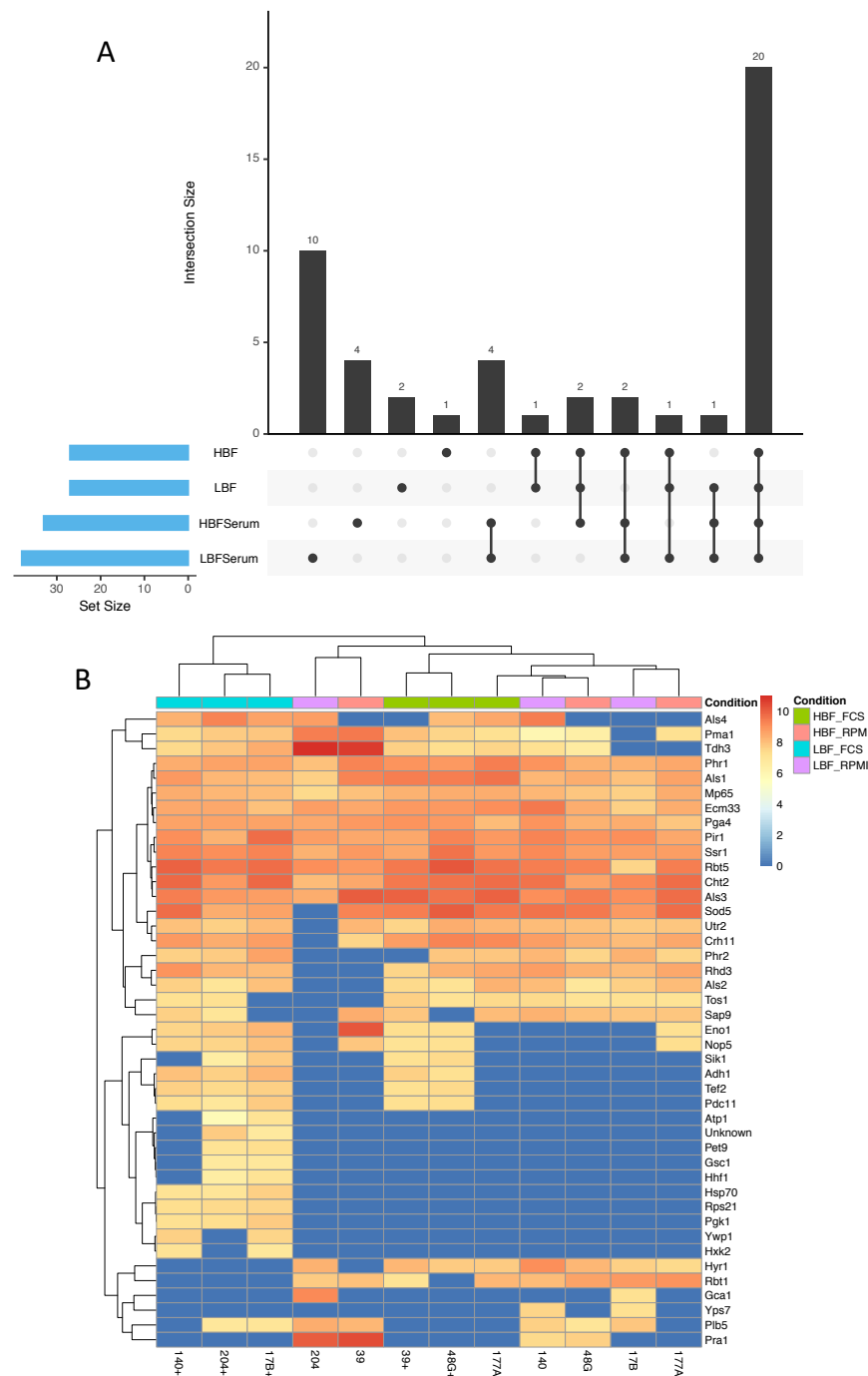
The Venn diagram like, Upset plot, shows the overlap of cell wall proteins identified between our different groups of HBF and LBF and in our two nutrient conditions (Figure 3.18A). Within our different groups 10 unique proteins were identified in the LBF serum group, 3 in the HBF serum group, 2 in the LBF no serum and 1 in HBF no serum group. The remainder of the proteins showed some level of overlap with one or more of the different groups with 20 proteins being extracted from all groups.

In the LBF strains 10 proteins were found in the cell walls in the presence of serum. These included the heat shock protein Hsp70, Essential beta-1,3-glucan synthase subunit Gsc1, the

small (40S) subunit Rps21, the ATP synthase alpha subunit Atp1, the secreted yeast wall protein Ywp1, the ATP carrier protein Pet9, the Phosphoglycerate kinase Pgk1, the Hexokinase II Hxk2, putative histone H4 Hhf1 and a protein of unknown function. The translation elongation factor Tef2, Pyruvate decarboxylase Pdc11, Alcohol dehydrogenase Adh1 and the mRNA binding and RNA processing Sik1 were identified only in serum but appeared in both the LBF and HBF strains. Interestingly the Enolase protein Eno1 was present in LBF in serum but not in the absence of serum. It is important to note that many of these proteins are mitochondrial proteins and therefore could cloud or results. Some may be due to moonlighting proteins however the likelihood is that they are present due to poor separation of cell wall during processing or due to contaminants from serum supplementation. Many of these potentially contaminant proteins however are consistently and uniquely in the LBF cells supplemented with FCS. This may be due to differences in the LBF cells or response to stress which have meant that extraction from this group was particularly difficult. This is not true of all LBF+FCS specific proteins however, the yeast wall protein YWP1 protein for example is a cell wall protein and is thought to be involved in the maintenance of biofilm formation in some instances despite its name indicating otherwise.

Between LBF and LBF+FCS there were 22 shared proteins, 5 only present in the LBF and 16 unique to the LBF+FCS. Between HBF and HBF+FCS there were 24 shared proteins, 3 that were unique to HBF and HBF+FCS. Between the HBF and LBF cells there were 24 proteins shared and 3 unique to HBF and 3 unique to LBF. Similarly, in HBF+FCS and LBF+FCS there were 27 proteins shared, 6 unique to HBF+FCS and 11 unique to the LBF+FCS. Shared proteins between all the phenotypes and media included ALS family proteins involved in adhesion between cells and surfaces. The GPI-Anchored Sap9 was also present across all phenotypes and nutrients and as previously discussed is consistently expressed in biofilm conditions. Cell surface mannoproteins are consistent across all biofilm conditions in the of Mp65 and the superoxide dismutase Sod5 which are involved in biofilm persistence (Bink et al., 2011). Most proteins seem to be typical cell wall proteins present in *C. albicans* and this can be observed by the overlap of the proteins in the Upset plot and the heatmap (Figure 3.18). Due to limitations with the methodology many of the atypical proteins observed could be erroneous due to poor extraction and or limited replication. Additionally, these results are summarised in the heatmap depicting the presence/absence and relative abundance indicated by log peak area

(Figure 3.18B). Presence/absence of identified proteins within each group and a description of their function are summarised in Table 3.5.



Chapter 3: Transcriptomic profiling of phenotypically distinct *Candida albicans* isolates

Table 3.5 Table of identified cell wall proteins and their presence/absence within each group of strains. Proteins also contain short description as maintained by the CGD. Proteins are indicated by presence (+) or absence (-) in each of groups. Groups are formed of the two phenotypes and the two nutrient conditions.

Names	Description	mycol	HBf	HBfSerum	LBF	LBFserum	Names	Description	mycol	HBf	HBfSerum	LBF	LBFserum
C1_00220W	Glycosidase	Phr2	+	+	+	+	C4_03520C	Cell wall protein with similarity to Hwp1 GPI-anchored yeast-associated cell wall protein	Rbt1	+	+	+	-
C1_01480C	Protein component of the small (40S) subunit	Rps21	-	-	-	+	C4_04050C	Cell surface glycosidase	Rhd3	+	+	+	+
C1_02420C	Essential beta-1,3-glucan synthase subunit	Gsc1	-	-	-	+	C4_04530C	Cell surface glycosidase	Phr1	+	+	+	+
C1_03190C	GPI-anchored cell wall protein	Ecm33	+	+	+	+	C4_04530C	Cell surface glycosidase	Phr1	+	+	+	+
C1_04240C	Putative histone H4	Hhf1	-	-	-	+	C4_06570C	Pyruvate decarboxylase	Pdc11	-	+	-	+
C1_04610W	ATP synthase alpha subunit	Atp1	-	-	-	+	C4_06980W	Cell surface protein that sequesters zinc from host tissue	Pra1	+	-	+	-
C1_08230C	Putative GPI-linked phospholipase B, fungal-specific (no mammalian homolog)	Plb5	+	-	+	+	C4_06980W	Cell surface protein that sequesters zinc from host tissue	Pra1	+	-	+	-
C1_08380W	Translation elongation factor 1-alpha	Tef2	-	+	-	+	C5_00590W	Mitochondrial ADP/ATP carrier protein involved in ATP biosynthesis	Pet9	-	-	-	+
C1_08500C	Enolase	Eno1	+	+	-	+	C5_04130C	GPI-linked chitinase	Cht2	+	+	+	+
C1_10290W	Extracellular/plasma membrane-associated glucoamylase	Gca1	-	-	+	-	C5_05050W	Alcohol dehydrogenase	Adh1	-	+	-	+
C1_13450W	GPI-anchored hyphal cell wall protein	Hyr1	+	+	+	-	C5_05050W	Alcohol dehydrogenase	Adh1	-	+	-	+
C1_13480W	Putative hsp70 chaperone	Hsp70	-	-	-	+	C5_05390C	GPI-anchored cell surface protein	Pga4	+	+	+	+
C2_00680C	Cu-containing superoxide dismutase	Sod5	+	+	+	+	C6_00370C	Ortholog of <i>S. cerevisiae</i> Nop58	Nop5	+	+	-	+
C2_02410W	Unknown	Unknown	-	-	-	+	C6_00750C	Phosphoglycerate kinase	Pgk1	-	-	-	+
C2_08590W	Secreted yeast wall protein	Ywp1	-	-	-	+	C6_03700W	Cell-surface adhesin	Als1	+	+	+	+
C2_08870C	1,3-beta-glucan-linked cell wall protein	Pir1	+	+	+	+	C6_04130C	GPI-anchored adhesin	Als4	-	+	+	+
C2_10030C	Cell surface mannoprotein	Mp65	+	+	+	+	C6_04380W	ALS family protein	Als2	+	+	+	+
C2_10030C	Cell surface mannoprotein	Mp65	+	+	+	+	C6_04380W	ALS family protein	Als2	+	+	+	+
C3_00720W	Plasma membrane H(+)-ATPase	Pma1	+	+	+	+	C7_00860W	Beta-glucan associated ser/thr rich cell-wall protein with a role in cell wall structure	Ssr1	+	+	+	+
C3_01550C	Protein similar to alpha agglutinin anchor subunit	Tos1	+	+	+	+	C7_02300W	Putative aspartic-type endopeptidase with limited ability to degrade alpha pheromone	Yps7	-	-	+	-
C3_01730C	Putative GPI anchored cell wall glycosidase	Utr2	+	+	+	+	CR_04510W	Hexokinase II	Hxk2	-	-	-	+
C3_03870C	Secreted aspartyl protease	Sap9	+	+	+	+	CR_04510W	Hexokinase II	Hxk2	-	-	-	+
C3_06870W	NAD-linked glyceraldehyde-3-phosphate dehydrogenase	Tdh3	+	+	+	+	CR_07070C	Cell wall adhesin	Als3	+	+	+	+
C4_00130W	GPI-linked cell wall protein	Rbt5	+	+	+	+	CR_09950C	Putative U3 snoRNP protein	Sik1	-	+	-	+
C4_02900C	GPI-anchored cell wall transglycosylase, putative ortholog of <i>S. cerevisiae</i> Crh1p	Crh11	+	+	+	+							

3.5 Discussion

Within clinical cohorts in which *C. albicans* is found, we have observed that there is variability within those populations of this yeast's biofilm forming ability. It has also been long understood that the biofilm forming phenotype is inducible. In experiments, primarily involving type and laboratory strains, it has been observed that numerous stimuli including nutrient sensing is able to invoke this phenotype (Kean et al., 2018a). This ability is not unique to nutrient stimulation and it has been discussed in greater depth within the introduction the numerous conditions that can moderate biofilm formation.

Through the work outlined here we observed that seemingly biofilm inefficient clinical strains are inducible to form a biofilm under stimulation with foetal calf serum. This effect of FCS outlined within the literature and in certain instances utilised to assure this phenotype within *in vitro* experiments (Chandra et al., 2001, Tammer et al., 2014). The use of FCS in *C. albicans* biofilm research has been leveraged to gain insights into the molecular and physiological changes that occur during biofilm formation (Pierce et al., 2015). Due to the clinical implications of biofilm formation and the consensus that biofilms result in a worse clinical outcome in many pathologies, then our increased understanding of the *C. albicans* biofilm allows for improving our knowledge of virulence and potential modes of interfering with the yeasts ability to harm *in vivo* (Cavalheiro and Teixeira, 2018).

We observed that FCS could induce biofilm formation into those strains that we had previously categorised as LBFs from the clinical cohort. Although the growth of *C. albicans* is sufficient to induce hyphal and biofilm growth in many isolates, it has been demonstrated that certain clinical isolates will more readily form biofilms than others. They took on the phenotypic properties visually of the HBF and their biomass and hyphal formation became much more comparable to the HBF strains. This effect was not seen in dialysed serum which has small molecules removed.

Transcriptomic profiling was hypothesised to identify similarities, deficiencies and contrasting changes that were occurring in the low biofilm forming strains. This allowed us to identify how the LBFs were regulating this phenotypic switching. RNA sequencing was utilised within

this research to identify changes in the transcriptional changes between two distinct clinical isolates with and without the nutrient stimulation of serum. RNA sequencing using the Illumina HiSeq platform was successful for all samples with all samples receiving adequate sequencing depth. There was a wide range of reads from ~11 to greater than 30 million for some of the samples. Average alignment rates were adequate as were the percentage of assigned genes. The Illumina platform provided adequate quality with no obvious failings within the data. The data was derived from single end reads which comparative to paired end reads do not allow for identification of novel features such as fusions or splice isoforms (Rossell et al., 2014). Single end reads are a simple and effective and high throughput method to achieve adequate read depth for broader RNA-Sequencing applications such as those utilised here. The identification and counting of features within our different conditions was our primary goal and this sequencing methodology was sufficient.

Both clinical *C. albicans* strains displayed a great deal of transcriptional plasticity when confronted with the FCS supplemented media. It would seem however that the degree of differentiation in gene expression is greater within the high biofilm forming. A larger number of genes were both up and down regulated in the HBF strain compared to the LBF strain. However, the largest determinant of transcriptional variability within our experimental samples was early and late biofilm formation. Which is perhaps not unexpected, variability due to strain and media changes was the 2nd and 3rd largest contribution to variability, respectively. Gene Ontology and over representation or enrichment analysis has become a popular tool for identifying patterns in gene expression (Conesa et al., 2016b).

Interestingly when considering our top differentially expressed genes between FCS and RPMI we observed that zinc transport was more upregulated in the RPMI only grown biofilms. ZRT1/2 zinc transport and PRA1 zincophore was consistently down regulated to a high level across all of our time points and biofilm conditions in FCS. The interaction between Pra and Zrt complexes allows for the efficient uptake of zinc in *C. albicans* (Łoboda and Rowińska-Żyrek, 2018). Zinc is required by many enzymes in yeasts and is thought to be a factor in *Candida* virulence and host interactions (Jung, 2015). Iron homeostasis has been linked to both fitness and biofilm formation in *C. albicans* however the intricacies of these factors and their roles are not fully understood (Mayer et al., 2013). Preliminary work by researchers in

Tokyo seemed to imply the increase of zinc in biofilm formation in *C. albicans* (Kurakado et al., 2018). Our data is slightly confounded by the HBF and LBF in RPMI where it is upregulated in both. HBF does form biofilms in RPMI and therefore may be biofilm related and LBF only form weak biofilms in RPMI, but it is still upregulated. Therefore, it is difficult to determine a role for zinc in biofilm formation with our dataset. It is however interesting to note that the supplementation of FCS is downregulating this response in both strains. This may indicate that zinc homeostasis is a key factor in biofilm formation and cellular elongation. Zinc maybe so readily available within the FCS that its metabolic needs are met earlier in biofilm formation. An even closer look at the initial stages of biofilm development and experimentation with increased zinc and zinc starvation in future experiments may help to decipher the role of zinc in biofilm formation.

From functional analysis it was possible to identify differentially programmed pathways and ontologies in the LBF in the presence of serum. Predictably, genes implicated in biofilm formation were changing at all time points. Cell adhesion and biofilm formation were also terms appearing in both HBF and LBF. Cell wall related terms, including fungal type and hyphal cell wall terms were represented at all time points and both HBF and LBF. Early biofilm formation in the LBF stimulated by the serum involved dioxygenase and preribosome. Late biofilm formation transcribed functional groups involved a larger number of processes. The largest majority was functional groups related to Acyl-CoA oxidase activity and fatty acid metabolism. β -oxidation of fatty acids in *C. albicans* takes place within the peroxisome. Acetyl units can then be integrated into the glyoxylate cycle or exported outside of the peroxisome by the carnitine shuttle system (Strijbis and Distel, 2010). These pathways are also overrepresented in the LBF in serum when compared to the HBF strains. Also distinct to LBF in serum compared to RPMI in the presence of serum is plasma membrane acetate transport. *C. albicans* utilising non-sugar carbon sources such as β -oxidation of fatty acids is known to be flexible and important for virulence in *C. albicans*. The fatty acid metabolites eicosanoids such as prostaglandin E2 and thromboxane B2, present in serum have been shown to influence germ tube formation in *C. albicans* (Noverr and Huffnagle, 2004). LBF strains can utilise alternative carbon metabolic pathways to enable yeast-hyphal morphogenesis and biofilm formation. We observed fatty acid metabolism and acetyl-CoA were distinctly LBF enriched terms in our low biofilm strains. The fatty acid binding and acyl-coA related genes

for Acyl-coenzyme A oxidase POX1 and POX 1-3 and PXP3 were specifically overrepresented in LBF. GSEA also revealed the fatty acid metabolism related genes FOX3, FOX2, ECI1, CAT2, FOX3, PEX11 and ANT1 were also upregulated and lead to the enrichment of the fatty acid oxidation related term. The fox 3-hydroxyacyl-CoA epimerase genes are involved in beta oxidation and induced in phagocytosis. PEX11 is a peroxisomal protein required for peroxisome proliferation and ANT1 is also described as a peroxisomal adenine nucleotide transported. EC1 and CAT2 are putatively linked to beta oxidation of medium chain fatty acids like the other genes. These genes are putatively assigned functions from the *S. cerevisiae* database with little annotation according to the STRING database.

Members of the ATO gene family ATO1, ATO10 and ATO9 were over-represented in the term “plasma membrane acetate transport” and they are upregulated in the LBF forming strains in the presence of FCS which are linked to acetate membrane transport and mutations in the ATO family have been demonstrated to affect the yeast to hyphal switch (Danhof and Lorenz, 2015). ATO genes are also thought to be dependent upon the transcription factor Stp2 (Danhof and Lorenz, 2015). Similarly, the associated genes FRP3 and FRP5 were upregulated associated with this pathway. FRP3 is a putative ammonium permease candidate. This pathway is associated with *C. albicans*' ability to induce alkalinisation through ammonia release and overcome macrophages (Danhof and Lorenz, 2015). They are involved in amino acid catabolic change and are induced to mitigate changes in pH. This need for alkalinisation may be a response from the LBF to account for changes in pH influenced by alternative carbon metabolism. pH is a known promoter of hyphal formation and is hyphal formation has been previously linked with alkalinisation (Vylkova et al., 2011a).

Iron homeostasis is upregulated at 4h in LBF in serum and downregulated at 24h. This could be indicative of *C. albicans* responding to higher levels of iron availability in serum compared to RPMI only. Iron concentrations previously being shown to instigate shifts from yeast to hyphal forms which previously has been theorised to be and an environmental trigger for *C. albicans* to reprogram its metabolism.

The *C. albicans* cell wall plays an important role in the cell morphology and protection against external insults. Adaptively, *C. albicans* remodels their cell walls in response to environmental

changes, a process controlled by well-known stress (Hog1) and cell integrity (Mkc1, Cek1) signalling pathways (Ene et al., 2015). The cell walls mainly comprised of β -glucans (glucose polymers [50-60%]), chitin (N-acetylglucosamine polymer [1-2%]), and mannoproteins (30-40%) (Poulain, 2015). Hsp70 has been evidenced previously in the cell wall of *C. albicans* indicating although not commonly found outside the cell membrane it has been hypothesised that they are molecular chaperones performing roles in cell wall biosynthesis (López-Ribot et al., 1996, López-Ribot and Chaffin, 1996). Hsp70 is responsible for protein folding under heat shock conditions and although it has been observed within the cell wall little is known about its function there. Gsc1 is an *S. cerevisiae* homolog where it is a subunit of the beta-1,3-glucan synthase (Mio et al., 1997) where it has a role in the production of beta-glucan a major component of the hyphal cell wall. Although similarly this is usually a membrane bound protein so its presence in the cell wall is unusual. The Rps21 ribosomal protein is localised within the endoplasmic reticulum but in the case of LBF in serum it was observed in the cell wall. Pet9 the ATP carrier is also localised within the mitochondria and not usually within the cell wall. Tef2, and Pdc11 are also not typically present on the cell wall. Pdc11 as well as several other proteins including Eno1 have previously been described as “moonlighting” proteins. Moonlighting proteins have been observed in response to chemical and host-interactions. It is perhaps only possible to say that LBF in serum is modulating the composition of its cell wall. We can only hypothesise that this is response to stress or compensating for deficiencies in cell wall biosynthesis comparative to HBF. This is an interesting question that cannot be answered within the work here but warrants further investigation. Additionally, as we discussed in the results the issues with cell wall protein extraction cannot be overlooked. Complications with proteomics methodology and preparation of cell wall fraction meant that potentially spurious results were introduced by presence of mitochondrial proteins. Although classified as an omics technology the cell wall proteome showed very little coverage and was not representative of the whole cell wall proteome. This again points to issues with extraction and processing that have been introduced through experimental design. It would be advantageous for this experiment to be repeated with greater number of replicates to both improve the power and to rule out any issues with contaminants. Results obtained were both limited and relative, so future experiments should aim to address the extraction of cell wall, processing to maintain a clean cell wall fraction and improve processing measurements to allow for quantitative measurements between sample groups.

It is fair to conclude that transcriptome analysis of our two strains has been successfully in determining that there is transcriptional reprogramming between these two strains. Biomarker analysis, however, found that their a select group of features that are consistent between the two strains and that can distinguish between RPMI and serum grown biofilms. With many transcriptional studies it drives hypothesis and creates as many questions as it answers. It is perhaps interesting that although LBF seem deficient, it is inducible through alternative pathways that we have shown here, and these are also pathways that have been observed within the host to induce hyphal morphogenesis such as the alkalinisation induced mechanisms within the macrophage. Therefore, these strains may still be capable of morphogenesis and invasion within the host.

3.6 Highlights

- *C. albicans* clinical isolates display heterogeneity in their ability to form biofilm but additionally in their morphological response to external stimuli.
- High and Low biofilm formers have drastically different gene expression profiles in response to foetal calf serum.
- The Low biofilm specific response to serum involves fatty acid metabolism and acyl-coA activity.
- Low biofilm forming isolates display potential “moonlighting” proteins in their cell wall not typically found.

4 Metabolomic profiling of phenotypically distinct *Candida albicans* clinical isolates

4.1 Background

The discipline of metabolomics continues to grow and so does its application to studying microbial biofilms (Burgain et al., 2020, Wong et al., 2018). It is becoming clear that primary and secondary metabolites play important roles in the biofilm formation of microbes, and therefore targeted and untargeted metabolomics have become an invaluable tool in the microbiologist's arsenal. Metabolomics offers a systematic and holistic method of identifying the metabolic profile during biofilm formation, and in turn the opportunity to identify important pathways and therapeutic targets. Recently metabolomics was used in *C. albicans* to identify all the key metabolic pathways that were reprogrammed in response to hypoxic stress and has previously been utilised to study metabolomic changes during morphogenesis and the metabolome in response to bacterial species such as *S. aureus* (Weidt et al., 2016, Burgain et al., 2020, Han et al., 2012).

C. albicans is a polymorphic fungus which can sense and adapt to its environment (Mayer et al., 2013). The "metabolic flexibility" has been linked and associated with *C. albicans* virulence and its interaction with host (Brown et al., 2014b). *C. albicans* and other fungal species possess mechanisms for metabolic adaption to amino acid and sugar metabolic pathways. Amino acid metabolism has been associated with the induction of hyphal morphogenesis through cAMP/PKA signalling pathways and through the alkalinisation of the environmental conditions (Silao et al., 2019, Vylkova et al., 2011b). Arginine metabolism in *Candida* spp. has been studied in some detail and is reported to induce *C. albicans* hyphal formation and elongation. This is linked to the cAMP-dependent protein kinase A (PKA) pathway and through CO₂ production thought to induce germ tube formation (Ghosh et al., 2009). Prior to interest in *C. albicans* as a biofilm forming opportunistic pathogen, it was well understood that morphogenic switching was observed in the presence of different nutrient sources, such as serum, amino acids (such as proline) and the monosaccharide N-acetylglucosamine (Feng et al., 1999, Dabrowa et al., 1976, Holmes and Shepherd, 1987).

Serum has been known for some time to be a potent inducer of hyphal morphogenesis and biofilm formation, through activation of the GTPase Ras1 dependent cAMP signalling leading to the phosphorylation of Efg1p (Ramage et al., 2002b). This leads to hyphal production in *C.*

albicans, and key structural component of biofilms. Hypoxia is another common stress and virulence response known to induce hyphal morphogenesis and biofilm formation. Metabolomic profiling by Burgain *et al* (2020) identified high levels of metabolomic reprogramming in *C. albicans* type strains in response to hypoxic stress (Burgain *et al.*, 2020). Importantly, early hypoxic stimulation leads to a response that causes a reprogramming of all major components of *C. albicans* metabolism, including carbohydrate, nucleotide, lipid and amino acid pathways. Similarly, it has previously been shown that sugar sensing or adaption to different sugar sources can influence the biofilm forming morphology of *C. albicans* (Pemmaraju *et al.*, 2016, Ng *et al.*, 2016). Availability of dietary carbohydrates can influence hyphal formation and biofilm development (Santana *et al.*, 2013). Most recently, through the use of heterogenous *C. albicans* clinical isolates it was shown that metabolic pathways including arginine, proline and pyruvate metabolism were transcriptionally upregulated in biofilm competent, or high biofilm forming isolates, compared to strains which were deficient in biofilm growth (Rajendran *et al.*, 2016a). This study indicated that amino acid metabolism was a key pathway supporting this phenotype.

Exploiting *C. albicans* metabolism has become an area of interest for therapeutic targets. It has been suggested that drugs could target and inhibit metabolic pathways, thereby reducing the virulence capacity of *C. albicans*. For example, through interruption of the glucose metabolism pathways in *C. albicans* through compounds that target the gluconeogenesis glyoxylate pathway. Recent examples include the uncompetitive inhibition of the trehalose-6-phosphate synthase and inhibition of the glucosamine-6-phosphate synthase inhibitors via N3-(4-methoxyfumaroyl)-(S)-2, 3-diaminopropanoic amide derivatives (Pawlak *et al.*, 2016, Magalhães *et al.*, 2017). Glycolysis has been the target of interest in a number of studies. Similarly, numerous studies have discussed potential, and in some cases, tested amino acid metabolism as a target for therapeutics because of their key roles in virulence and morphogenesis (Garbe and Vylkova, 2019). Similarly, reports exist on effects of attenuation of amino acid pathways in other fungal species such as *Cryptococcus neoformans*, *Aspergillus fumigatus*, as well as for *C. albicans* (McCarthy and Walsh, 2018, Jastrzębowska and Gabriel, 2015). For example, in *C. neoformans* disruptions of the catabolic proline pathway was observed to decrease virulence through reducing the organism's ability to respond to oxidative and nitrosative stress (Lee *et al.*, 2013). In *A. fumigatus* it has also been proposed

that due to this dependence on amino acids during aspergillosis. As the disruption of carbon metabolism linked to amino acid metabolism such as methycitrate cycle can attenuate virulence (Ibrahim-Granet et al., 2008).

4.2 Aims

We hypothesised that using targeted/untargeted metabolomics that we could identify metabolomic differences between *C. albicans* phenotypically distinct isolates defined as high and low biofilm formers. Additionally, we hypothesised that we could identify metabolic changes that occur during the phenotypic shifting of *C. albicans* in the presence of serum. This chapter aims to interrogate the differentially modulated pathways under the biofilm inducing condition of serum supplementation.

We additionally aimed to identify any metabolic response or reprogramming in *C. albicans* low, or biofilm deficient, clinical isolates compared to high or biofilm capable isolates.

Specifically, this chapter aims to identify:

- Differences in the metabolomic footprint of key metabolites using targeted LC-MS metabolomics.
- Difference in the targeted metabolomic footprint which was unique to the low or high biofilm forming clinical isolates
- Pathways that were reprogrammed or underwent changes in the response to serum using untargeted metabolomics
- Pathway activity which was unique or specific to the either the low or high biofilm clinical isolates.

This work has been presented at the following conferences:

- Eurobiofilms, Amsterdam 2017 - **Integration of Metabolomics and Transcriptomics in Biofilm Research (Poster, Recipient of a travel grant to attend from the organisers)**

- ECCMID, Madrid 2018- ***De novo* transcriptome assembly and annotation for analysis of the emerging pathogen *Candida auris* (Poster)**
- OMIG, Gregynog 2018- **Omics approaches to studying microbial biofilms in oral health (Talk, Recipient of a travel grant to attend from the organisers)**
- IADR, London 2018- **Omics and bioinformatics approaches to studying oral biofilms (Poster)**
- Eurobiofilms, Glasgow 2019 - **Transcriptomic and multi-omics data integration approaches to interrogate mono and interkingdom species *Candida* biofilms (Talk)**
- Invited talk, Newcastle 2020 - **OMIC modelling of *Candida albicans* interkingdom biofilm interactions (Seminar)**

4.3 Methods

4.3.1 Biofilm Preparation

HBF (n=3) and LBF (n=3) isolates were grown as biofilms for 4 and 24h in RPMI ± FCS, as described previously, in triplicate for a total of 9 biofilms per sample group. All isolates used in this component of the thesis are summarised in Table 3.1 (found in section 3.3.3).

Biofilms were grown as previously stated in polystyrene plates at 30°C. 5µL of the spent media was added to 200µL of an ice cold solvent solution containing chloroform:methanol:water in a (1:3:1:v:v:v) ratio. This solution was then vortexed at maximum speed for 15sec and stored at -80°C. The mixture solution was separated using Hydrophilic interaction liquid chromatography on SeQuant® ZIC® -pHILIC columns (Merck KGaA, Darmstadt, Germany), before LC-MS analysis performed on the Thermo Scientific Exactive Orbitrap system (Thermo Scientific, Hemel Hempstead, UK). Fresh media harvested at both 4h and 24h was used as a control and solvent solution was used as a blank. Thermo Xcalibur™ (version 2.2.42) was used for instrument control and data acquisition.

4.3.2 Data Acquisition and Processing

Mass spectrometry raw data files were filtered and converted to mzXML file format using the msconvert component of the ProteoWizard software package (Chambers et al., 2012). mzXML files were submitted to the Polyomics integrated Metabolomics Pipeline (PiMP) (<http://polyomics.mvls.gla.ac.uk/>) (Gloaguen et al., 2017). In summary the integrated pipeline utilises the program XCMS for peak picking and MZmatch for peak matching, noise filtering and annotation of related peaks. Principal metabolites were identified against a panel of authentic pure standards (Sigma-Aldrich), that are maintained in the University of Glasgow Polyomics standard compound library, using accurate mass and RT (error 5%) and were defined as Identified based upon the Metabolite Standards Initiative. Putatively assigned compounds using the accurate mass and to give a predicted compound.

4.3.3 Statistical Analysis

Once peaks are identified their intensities was logged by base 2 and pairwise analysis was performed on group-to-group comparisons. PiMP is a web-based interface which allows for the uploading, hosting, processing, and analysis of metabolomics datasets. It includes pairwise and functional analysis parameters processed data was exported as abundance tables for further analysis.

Tables were exported with peak Ids and the corresponding KEGG and compound identifiers, in addition to the identification status against the standard compounds. A total of 612 peaks were present within the data, of these, 49 corresponded to the known standards and the rest were identified putatively against the KEGG database compounds. Principal Component Analysis was performed on total sum scaled and log transformed peak relative abundance data to visualise overall variance within the data. The relative abundance for each sample was normalised to the relevant media controls before principal component analysis (PCA) to visualise variance after normalisation to the control media. Analysis of variance (ANOVA) was performed on the identified compounds followed by Benjamini-Hochberg false detection rate FDR of the p-values to account for the numerous ANOVA tests performed. Those metabolites which were still deemed to be significant following FDR of p-values were selected for further analysis and the post hoc Tukey test was performed. Tukey tests determined the significance between RPMI and Serum for each of the metabolites.

With many mass peaks being putatively assigned compounds accurate detection of all compounds in untargeted metabolomics is not possible. Therefore, compounds were analysed according to their contribution to metabolic pathways. Activity scores of the KEGG functional groups were calculated for each of the samples using the R package Pathway Activity Profiling algorithm (PAPi)(v1.12.0) (Aggio et al., 2010). ANOVA followed by Tukey post hoc test were performed to determine significantly changing pathways between our sample groups within base R. Partial Least Squares Discriminant Analysis was performed on all identified peaks for biomarker analysis. PLS-DA. PLS-DA is a multivariate supervised analysis useful in class discrimination. It is a classification approach which is also able to determine discriminatory markers between our groups. Important features are discernible through their variable importance in the projection (VIP) scores. PLS-DA was performed within the MetaboAnalyst (V4.0) software utility following the data being exported sample and peak relative abundance tables (Chong et al., 2018a). Biofilm processing data acquisition and analysis is summarised in Figure 4.1.

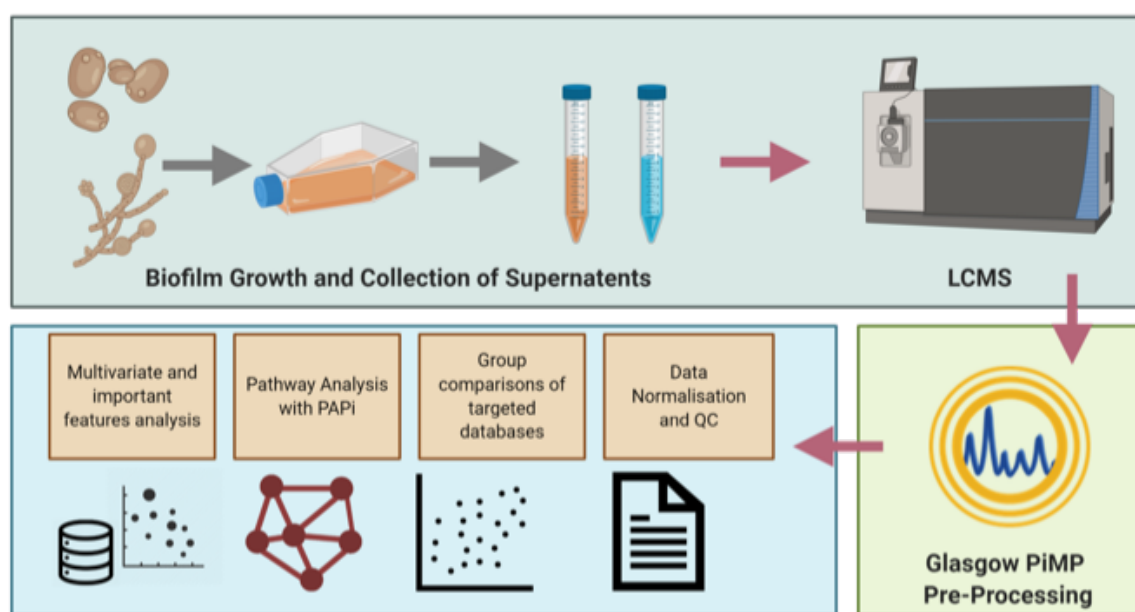


Figure 4.1 **Schematic depicting our workflow.** This diagram illustrates from the wet lab stage of biofilm growth of biofilms and collection of supernatants to mass spectrometry and the pre-processing by the Glasgow Polyomics in house analysis server and then final data analysis and interpretation through R statistical programming analysis and online analysis databases.

4.4 Results

4.4.1 Comparisons between media controls

Volcano plots, scatter plot of log₂ fold change (log₂FC) against the negative log₁₀ adjusted p-value, were created for the comparison of media ± FCS after 4h and 24h incubation. A positive fold change indicates that the peak was observed at a higher relative abundance in the media ± FCS. 16 different peaks were observed at the 24h time point and 20 peaks at the 4h time point (Figure 4.2). 8 of these metabolites were Identified within the standards and were higher in the media ± FCS. In RPMI + FCS compared to RPMI the identified metabolites in order of the significant abundance were Creatinine (p=3.8e-09), L-Glutamine (p=1.2e-07), Ornithine (p=0.00023), Hypoxanthine (p=1.1e-05), L-Malic acid (p=0.0022), Pantothenate (p= 0.013), Choline (p=0.016) and O-Acylcarnitine (p= 0.0051).

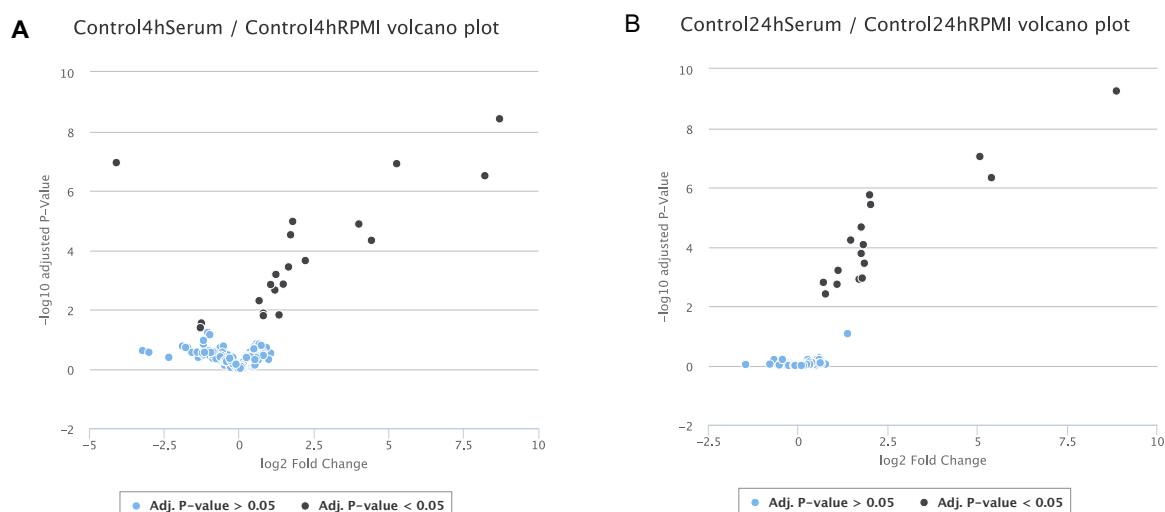


Figure 4.2 Volcano plots depicting differentially represented metabolites in our media controls. Metabolite abundance was compared between the media controls RPMI and RPMI supplemented with FCS by LCMS. This was processed within PiMP to derive significance and log₂ fold change between the two medias. Media that had been incubated for 4h and 24h was compared. The volcano plots depict the features changing between RPMI and RPMI supplemented with FCS. Comparison is between the media controls at 4h (A) and 24h (B).

The most significantly changing metabolites after FDR are displayed with a log₂ FC >2 are shown in Table 4.1. Most of these metabolites were involved in amino acid metabolism followed by carbohydrate and energy metabolism related metabolites.

Table 4.1 Fold changes of FCS metabolites compared to RPMI. Metabolite abundance was compared between the media controls RPMI and RPMI supplemented with FCS by LCMS. This was processed within PiMP to derive significance and log2 fold change between the two medias. This table depicts those metabolites that were different between RPMI and RPMI supplemented with serum after 4 and 24h. with a fold change greater than 1.5 indicated by light shading and dark shading of red (+ve) and green (-ve) fold change greater than 2.

Metabolite	Formula	logFC 4hFCS/4hRPMI	logFC 24hFCS/24hRPMI	Identification
Creatinine	C4H7N3O	8.71	8.88	I
3-(N-Nitrosomethylamino)propionitrile	C4H7N3O	8.71	8.88	A
Creatine	C4H9N3O2	8.23	5.4	A
3-Guanidinopropanoate	C4H9N3O2	8.23	5.4	A
L-Glutamine	C5H10N2O3	5.26	5.07	I
Alanylglycine	C5H10N2O3	5.26	5.07	A
3-Ureidoisobutyrate	C5H10N2O3	5.26	5.07	A
D-Glutamine	C5H10N2O3	5.26	5.07	A
Glutamine	C5H10N2O3	5.26	5.07	A
Isoglutamine	C5H10N2O3	5.26	5.07	A
Mannitol	C6H14O6	4.01	1.7	A
Galactitol	C6H14O6	4.01	1.7	A
D-Sorbitol	C6H14O6	4.01	1.7	A
L-Iditol	C6H14O6	4.01	1.7	A
L-Glucitol	C6H14O6	4.01	1.7	A
D-Iditol	C6H14O6	4.01	1.7	A
L-Ornithine	C5H12N2O2	2.2	1.85	I
D-Ornithine	C5H12N2O2	2.2	1.85	A
2,5-Diaminopentanoic acid	C5H12N2O2	2.2	1.85	A
(2R,4S)-2,4-Diaminopentanoate	C5H12N2O2	2.2	1.85	A
Hypoxanthine	C5H4N4O	1.79	2	I
Allopurinol	C5H4N4O	1.79	2	A
N-Acetylhistidine	C8H11N3O3	1.72	2.01	A
HC Red No. 3	C8H11N3O3	1.72	2.01	A
N-Acetyl-L-histidine	C8H11N3O3	1.72	2.01	A
(R)-Piperazine-2-carboxamide	C5H11N3O	-3.22	-0.06	A
(S)-Piperazine-2-carboxamide	C5H11N3O	-3.22	-0.06	A
4-Guanidinobutanol	C5H11N3O	-3.22	-0.06	A
Dimethyl sulfoxide	C2H6OS	-2.35	-1.45	A
Mercaptoethanol	C2H6OS	-2.35	-1.45	A

RPMI media is comprised of sugars (glucose and glutathione), pH indicator (phenol red), inorganic salts (sodium chloride, sodium bicarbonate, disodium phosphate, potassium chloride, magnesium sulfate, and calcium nitrate), vitamins *i*-inositol; choline chloride; *para*-aminobenzoic acid, folic acid, nicotinamide, pyridoxine hydrochloride, and thiamine hydrochloride; calcium pantothenate; biotin and riboflavin) and amino acids (glutamine; arginine; asparagine, cystine, leucine, and isoleucine; lysine hydrochloride; serine; aspartic acid, glutamic acid, hydroxyproline, proline, threonine, tyrosine, and valine; each histidine, methionine, and phenylalanine; glycine; tryptophan; and reduced glutathione). RPMI is a well-defined media typically produced for cell culture and has been commonly utilised for

culturing *C. albicans*. Serum or FCS however is undefined media with uncontrolled levels of constituents. However, it is a mixture of sugars such as glucose, hormones such as insulin, testosterone and progesterone, creatinine, bilirubin, proteins, salts, vitamins A/B and prostaglandins.

4.4.2 Metabolomic profiling of Low and High Biofilm Forming Isolates

The initial comparison exists between the low and high biofilm formers in RPMI with comparisons to changes that exist in those that are grown in Serum. We have previously identified transcriptomic changes between high and low biofilm formers. Between LBF and HBF grown for 24h in RPMI we identified significant changes (adjusted p-value less than 0.05) of 10 identified compounds and 19 annotated compounds. We identified 2 annotated compounds significantly changing in between HBF and LBF. Comparatively we identified 7 identified compounds and 12 annotated peaks significantly changing at 24h between HBF and LBF in serum. This was much less at 4h in serum with 2 identified compounds significantly changing and 7 annotated peaks meeting the significance criteria. Volcano plots depicting these changes at each time point and media are shown in Figure 4.3. Blue dots indicate peaks that were not significant where those coloured black indicate that they were significant.

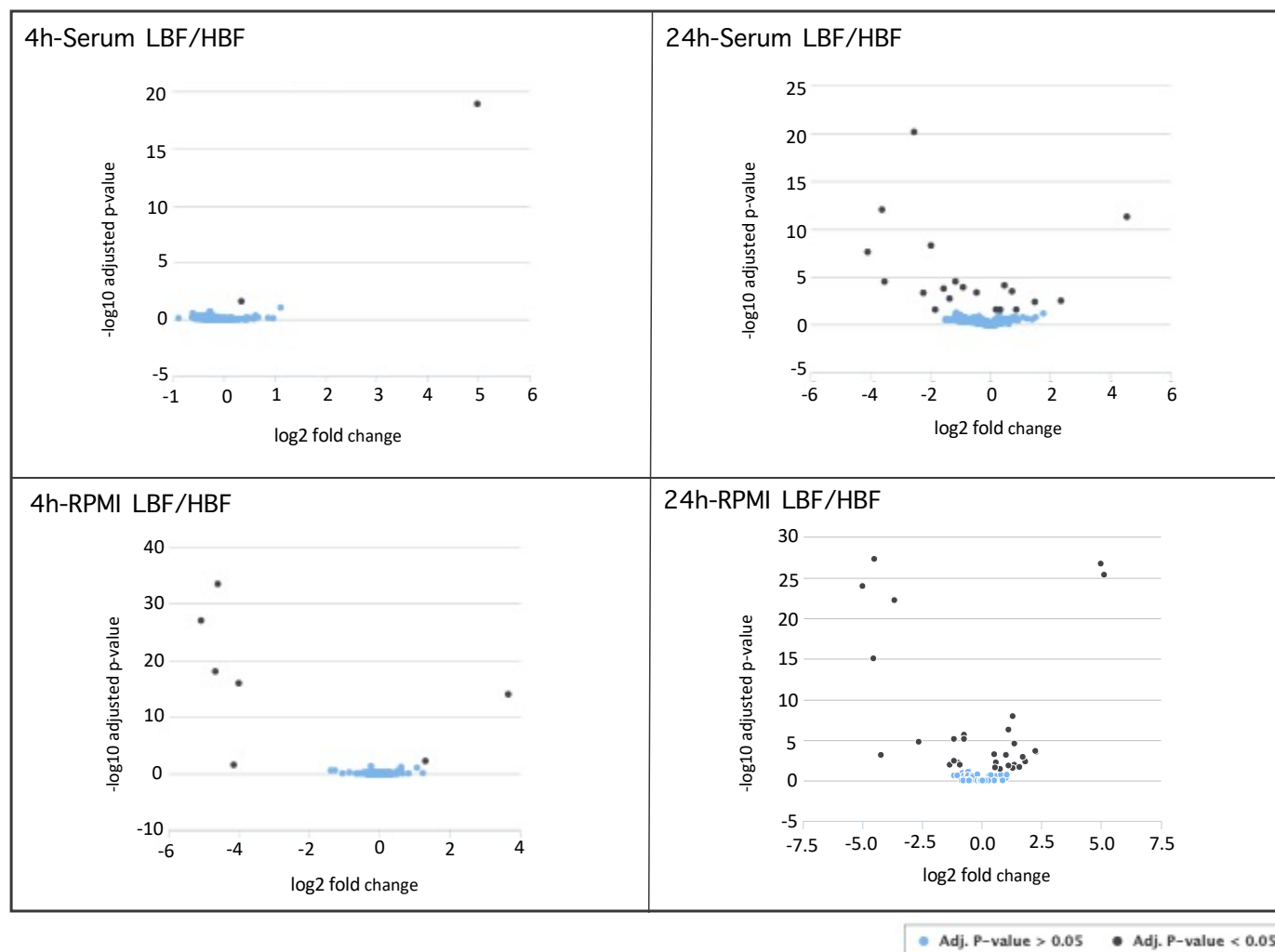


Figure 4.3 **Volcano plots showing differences between LBF and HBF at each respective timepoint in both RPMI and RPMI+FCS.** Metabolite abundance was compared between high (HBF) and low (LBF) biofilm forming strains at 4h and 24h in both RPMI and Serum. This was processed within PiMP to derive significance and \log_2 fold change between the two strains. Significantly different metabolites between HBF and LBF are indicated with black dots. Significance is indicated by negative \log_{10} of the adjusted on they y axis and \log_2 fold change is indicated on the x.

Between the LBF and HBF in RPMI, the most significantly changing metabolites in higher relative abundance in the low biofilm forming supernatants were L-Aspartate, L-phenylalanine, L-glutamate, L-glutamine, L-tyrosine and Hydroxyproline trans-4-Hydroxy-L-proline according to the adjusted p-value <0.05. The most significant differences in serum according between the low and high biofilm formers were Betaine, sn-glycero-3-Phosphocholin which were in higher relative abundance in the high biofilm forming strains media. L-glutamate, L-Aspartate and L-Arginine and O-Acetyl carnitine were in higher abundance in the media of the LBF (Figure 4.4).

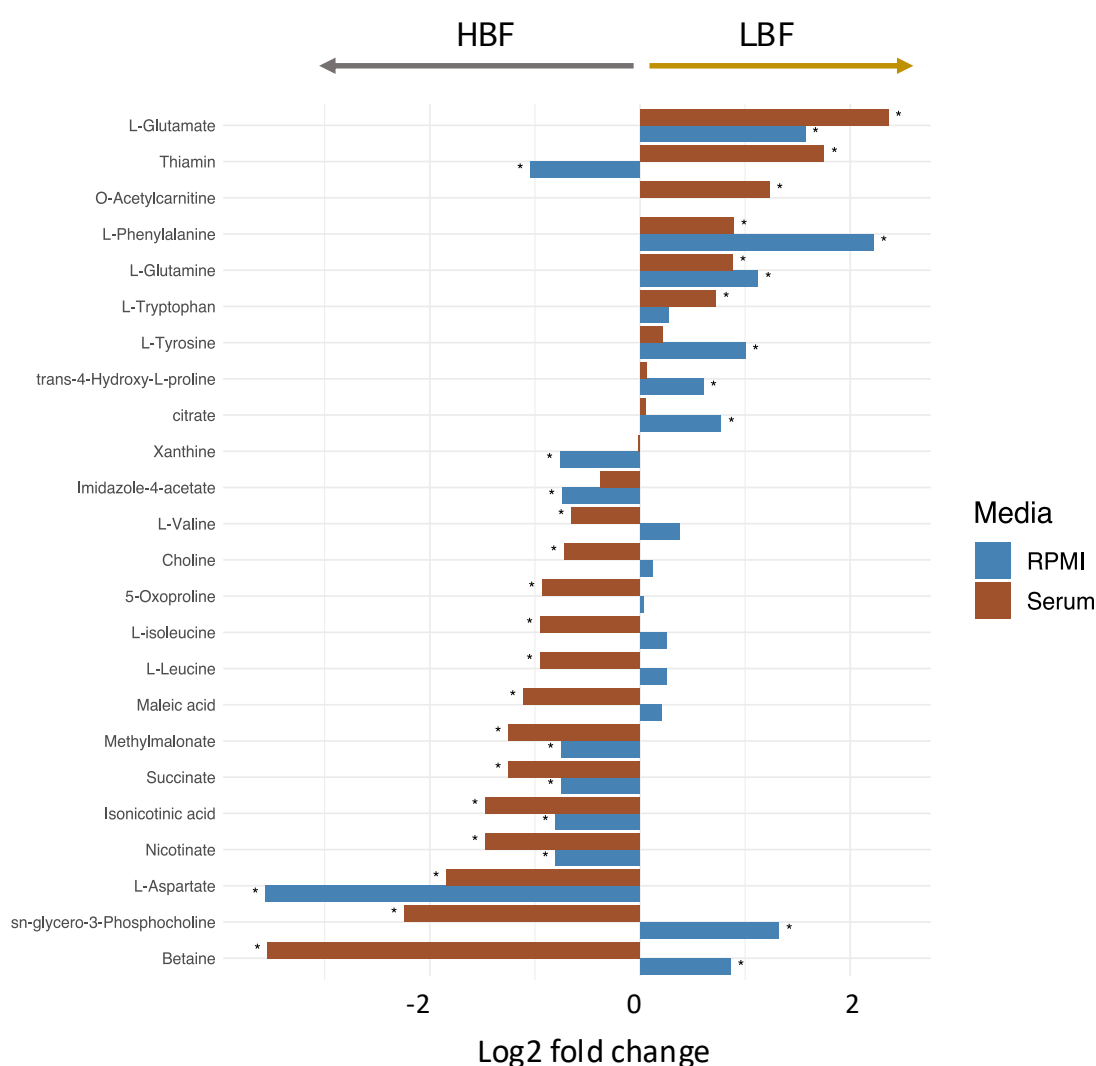


Figure 4.4 Bar chart of significantly changing identified compounds between HBF and LBF. From the analysis within the PiMP platform univariate comparisons between HBF and LBF was performed in for both Serum and RPMI independently. The log2 fold change and their significance (p<0.05) was then exported. Metabolite fold change between HBF and LBF grown for 24h in RPMI are coloured red and those grown in serum are coloured blue. Significant p-value of less than 0.05 than within each of the pairwise conditions HBF/LBF in their respective mediums is indicated by (*). A +ve log2 fold change indicates a higher level in the LBF strains and -ve log2 fold change indicates a higher level in the HBF.

Through Pathway Activity Level Scoring within PiMP, it was also noted that in order the top five pathways significantly changing were the bacterial chemotaxis, Lysine biosynthesis, beta-Alanine and metabolism, the Glutamatergic synapse and Nitrogen metabolism where the most enriched KEGG pathways in RPMI at 24h between HBF and LBF. In serum at 24h the top five significantly changing where, in order, sulphur metabolism, glutamatergic synapse, FoxO signalling pathway, bacterial chemotaxis and nitrogen metabolism.

The FoxO pathway is associated the identified or annotated compounds D-glucose and L-glutamate and was significant in serum between LBF and HBF due to the higher levels of glutamate in the HBF media. The FoxO pathway is a mammalian signalling pathway not typically found in microorganisms.

Bacterial chemotaxis is a KEGG pathway which is not specific to *C. albicans* but is enriched by the identified L-aspartate. This is due to a limitation within the PiMP software and it being unable to filter the KEGG pathway assignment based upon species. The glutamatergic synapse was enriched between LBF and HBF due to the high peak abundance of L-glutamate followed by L-glutamine. Both were in higher abundance in the HBF strains media after 24h. Sulphur metabolism was enriched due to the significantly higher levels of the p-value 0.05 of O-Acetyl-L-Serine in the LBF media and the significantly higher levels of Succinate in the HBF strains media. Nitrogen metabolism is similarly enriched by the L-glutamate and L-glutamine compounds. Many of the pathways within the pathway analysis included in the PiMP processing pipeline are unlikely to be representative of *C. albicans*. Currently the pathway profiling within the PiMP platform relies upon the larger, more general, KEGG database rather than species specific ones. As such many of the pathways are not specific to the organism being studied. In addition, because of the small differences in metabolite abundance between HBF and LBF the pathways that are overrepresented based upon small numbers of metabolites.

PLS-DA shows the top discriminatory compounds between HBF isolate and LBF (Figure 4.5A). This comparison shows the direct changes of metabolome directly related to the two different strain types. L-phenylalanine, L-glutamate and L-tryptophan distinguish between the low and high biofilm formers according to the variables importance in the projection (VIP) (Figure

4.5B, C). Additional peaks also distinguish between the two isolate types however they show decreasing importance in the projection and due to being putatively identified these compounds could be a number of different metabolites. The top discriminatory variables were also observed in our univariate analysis described above.

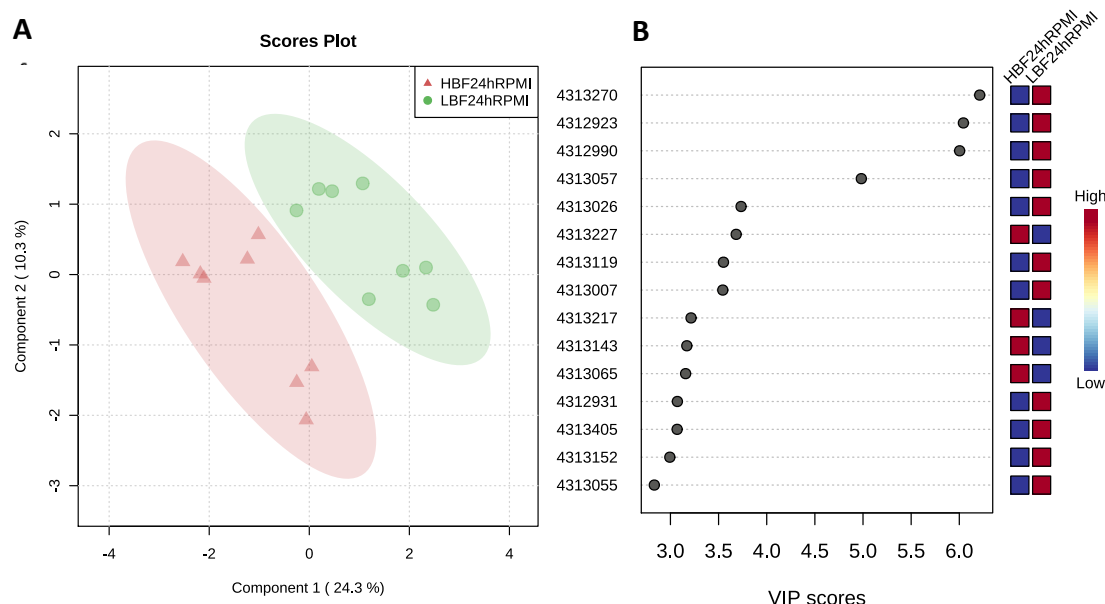


Figure 4.5 **Partial Least Squares of LBF and HBF *C. albicans* isolates at 24h.** The metabolite abundance values identified by the PiMP software were submitted to MetaboAnalyst for partial least squares analysis (PLS). PLS was performed for Identification of the most discriminatory metabolites between HBF and LBF. HBF (red) are distinguished here from the LBF (green) (A). Variable importance in the projection (VIP) scores of the top discriminatory compound peaks (B). These scores indicate how impactful they are at discriminating the two conditions in order of importance. These compound peaks were then compared to their compound or metabolite name if they could be identified. Identification status and Name or compound of the VIP peaks are described in a table (C).

4.4.3 Multivariate analysis of *C. albicans* HBF and LBF

Principal component analysis (PCA) displayed good separation of the different groups between our two medias and time points (Figure 4.6A). However, little separation was observable between our two different strain phenotypes (i.e., HBF and LBF). Samples also clustered together in groups of media and time. We have previously shown that there are strong phenotypic differences and previously identified metabolomic profile within the transcriptome (Rajendran et al., 2016a). Within our study design of samples strain to strain variation did not account for much of the total variation observed as shown with the PCA. For example, the samples grown for 24h in RPMI formed a distinct cluster. Data was normalised to media controls, by dividing the metabolite abundance by of the sample by the metabolite abundance within the media only controls, the values became log relative compared to the media control. This alleviated the high skew within the data and removed some of the heteroscedastic within the data (Figure 4.6B). It also improves the interpretability when comparing the footprint of the organisms between two different medias. As some values will be positive, indicating that they are added or secreted to the media by the cells, or the values would become negative indicating that they had been taken away from the media or spent. PCA projection of the normalised data again showed separation of the data by time and by media. Similarly, to the non-media normalised data our samples seemed to group and cluster according to their incubation times and media. One outlier was also removed from the data according to a cut of 50% or more missing features for further analysis. Other failures prior to processing within PiMP, due to failures during mass spectrometry and/or data acquisition meant that only two sample groups had a total of 9 replicates. However, the lowest number of replicates per sample was 6 during PiMP processing and this was deemed adequate. Samples also showed clear reproducibility within their metabolomic profiles (Figure 4.6). All peaks and successful samples and their corresponding peaks are available in additional appendix I.

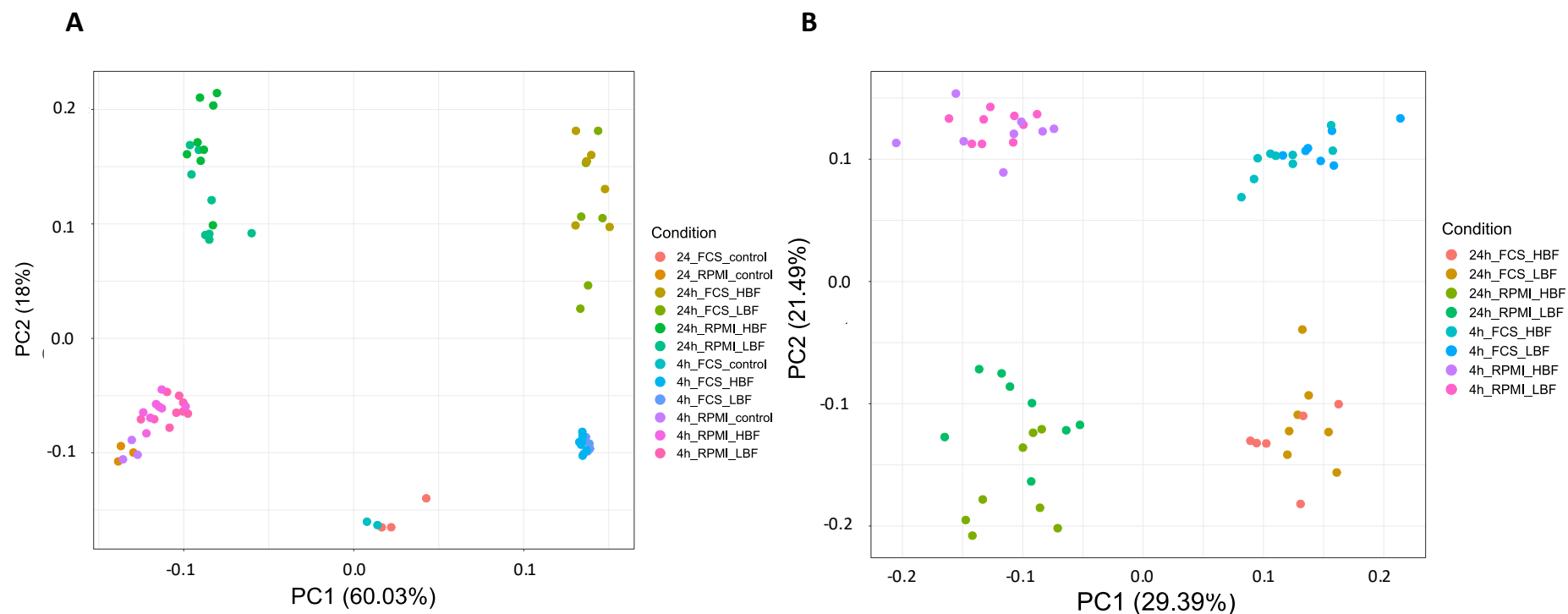


Figure 4.6 **Principal component analysis (PCA) of metabolomics secretome/metabolic footprint analysis of *C. albicans* high and low biofilms in the presence or absence of serum at 4h or 24h.** All untargeted metabolites were included in the principal component for all samples (A). The highest variability within the sample set is shown on PC1 (60.03%) with the second highest shown on PC2 (18%) (A). PCA of the same metabolomics data normalised to the growth media controls with the largest variation between the media corrected samples PC1 (29.39%) and the second largest variation on component 2 (21.49%) (B). Colourisation of the points is according to the sample group. Sample group is indicated by biofilm growth time_media_strain.

4.4.4 Profiling of targeted metabolites in isolates grown in serum vs RPMI

Within the different samples a total 49 compounds matched to a known standard from the Polyomics standards collection. Many metabolites within the standard compound library were not observed. These may be due to metabolites being below the detection limit of the machinery. More plausibly it could be due to the limitations of doing metabolic foot printing or extracellular metabolomics. Many intermediate metabolites within the metabolic networks of *C. albicans* may not have been externalised during growth and biofilm formation. However, this limitation is consistent throughout our testing and therefore comparison of the metabolic footprint left between our different sample groups still allows comparison and inference of the metabolic pathways being differentially regulated between our samples. Of the 49 metabolites that were identified, 38 were found to be in significantly different abundance between our groups. Of these, 21 had a difference in abundance between the sample groups according to ANOVA following correction for multiple tests with an FDR p-value. Identified metabolites were grouped into super pathways according to their KEGG designation for visualisation. Although, many of the metabolites appear in several different KEGG database pathways. The 49 compounds relative fold change compared the media is summarised in Figure 4.7. Additionally, their super pathway grouping is also shown. Due to the data being a log₂ fold change relative to the media control an increase or positive fold change indicates an increase in the media (the metabolite is secreted or released) and the larger the negative value indicates that the metabolite is spent relative to the negative control. ANOVA was performed for each of the metabolites, and they were considered significant after these values were corrected if the FDR p-value ≤ 0.05 . Significant differences between RPMI and FCS grown strains was determined using the Tukey's honestly significant difference (HSD) ≤ 0.05 and are discussed in the following sections.

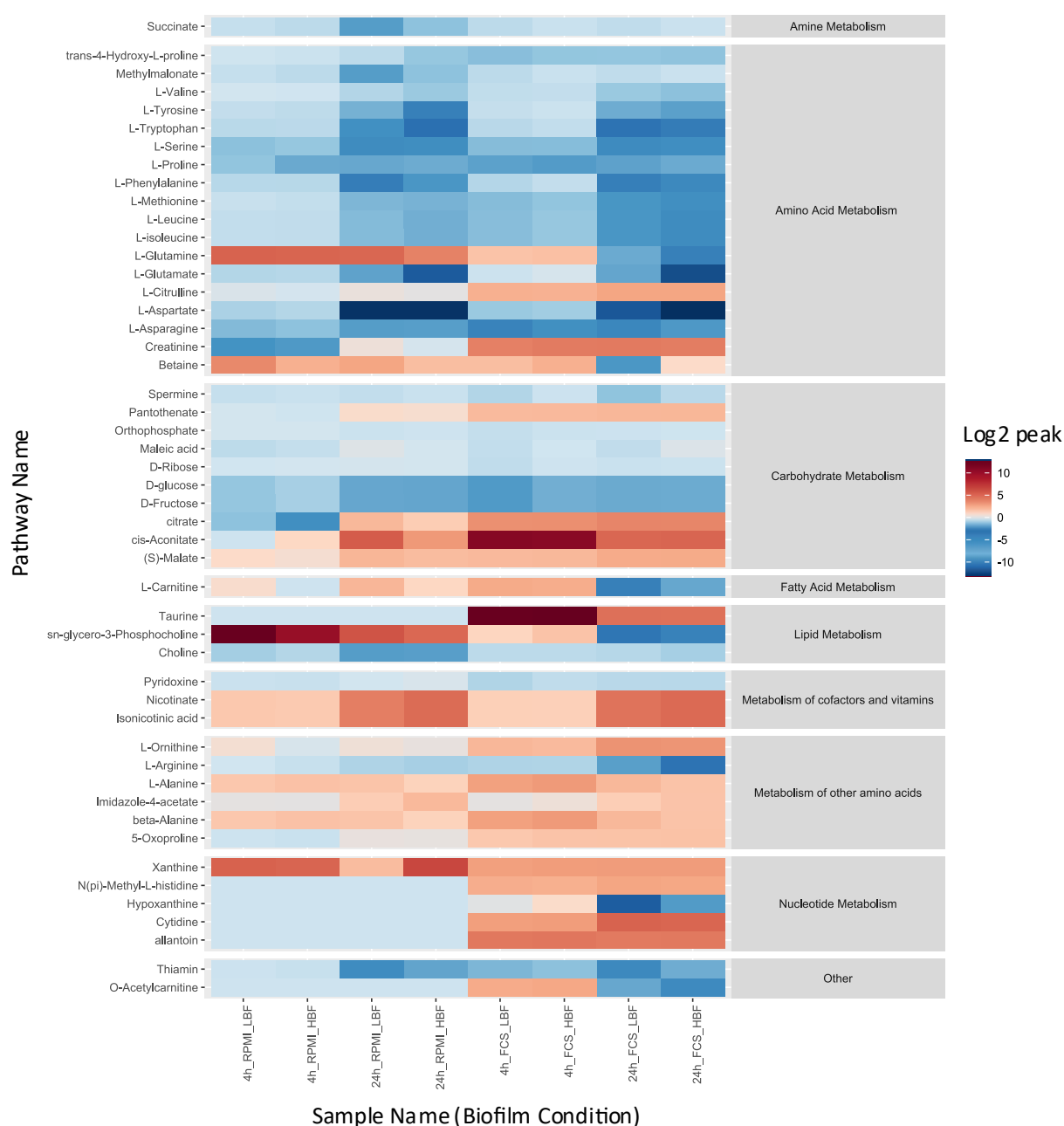


Figure 4.7 Grid or heat plot of the relative abundance compared to the media control of the targeted or identified metabolites. Positive log2 values infer secretion of metabolites to the media and negative values infer that metabolites have been spent from the media. Values are log2 metabolites relative abundance which have been divided by the control values. The values have also been grouped into their super-pathway metabolic groups according to their KEGG classification. These values are grouped into each sample group which is determined as time_media_strain.

4.4.5 Identified amino acid metabolism

Many of the identified metabolites are amino acids (24 identified metabolites), and of these 11 were found to be in significantly greater abundance between RPMI and RPMI+FCS as determined by Tukey HSD <0.05 (Figures 4.8-4.9). Betaine is a metabolite in *C. albicans* that is found in the threonine, serine and cysteine metabolic pathways. Betaine was observed to be spent in FCS compared to RPMI in our LBF strain at 24h ($p < 0.0001$). A difference in abundance is observable between FCS and RPMI in the HBF strain at 24h, with the FCS media showing less betaine secreted. However, this was not found to be significant. L-glutamine is a metabolite found in *C. albicans* in the arginine, purine and pyrimidine metabolism, as well as several other different metabolic pathways. It was observed to be excreted at a lower level in the 4h time points for both LBF and HBF in FCS and spent in comparison to excreted at 24h. This was significant for both strains at both time points with a $p < 0.0001$ (Figure 4.8).

Amino Acid

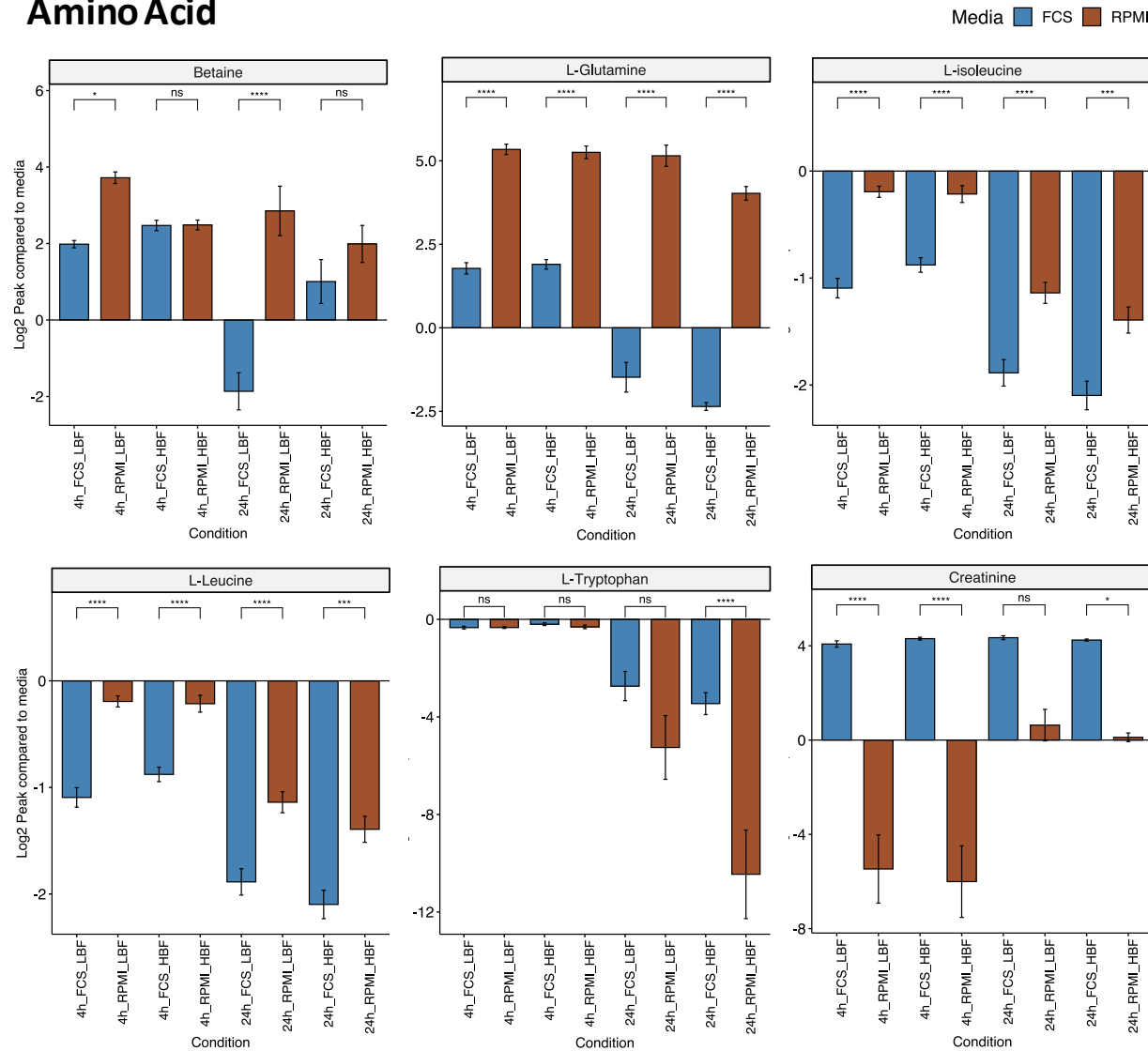


Figure 4.8 Identified metabolites between RPMI and Serum in *C. albicans* isolates. Bar charts depicting the significantly different amino acid related metabolites selected by Analysis of Variance (ANOVA) followed by adjusting the p-value using Benjamini-Hochberg FDR. Identification of significant changes was performed using the Tukey HSD. Values are log2 peak area compared to media control with positive values indicating secretion and negative values indicating that the metabolite was spent compared to media control. For each plot significance is indicated according by the Tukey derived p-value where there is a significant difference between the log2 relative area in RPMI vs Serum in either HBF or LBF and at either 4 or 24h. Significance is indicated between the four comparisons as ns, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** or $p < 0.0001$ ****.

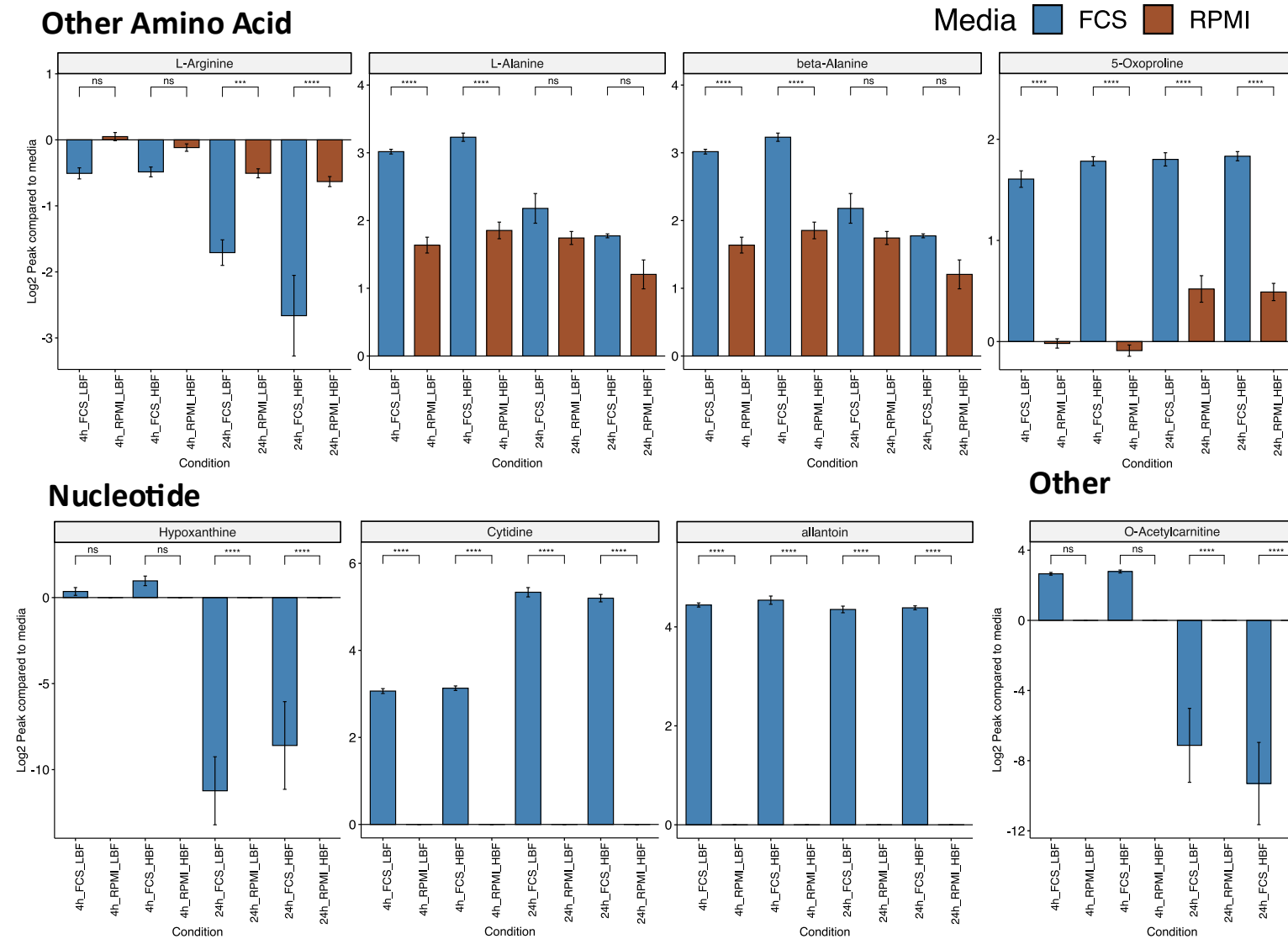


Figure 4.9 **Identified metabolites between RPMI and Serum in *C. albicans* isolates.** Bar charts depicting the significantly different amino acid and nucleotide related metabolites selected by Analysis of Variance (ANOVA) followed by adjusting the p-value using Benjamini-Hochberg FDR. Identification of significant changes was performed using the Tukey HSD. Values are log2 peak area compared to media control with positive values indicating secretion and negative values indicating that the metabolite was spent compared to media control. For each plot significance is indicated according by the Tukey derived p-value where there is a significant difference between the log2 relative area in RPMI vs Serum in either HBF or LBF and at either 4 or 24h. Significance is indicated as ns, p<0.05*, p<0.01**, p<0.001*** or p<0.0001****.

L-isoleucine was found to be spent to a higher degree in FCS at all time points and in both strains. It was found to be significant between serum and RPMI at 4h in LBF($p=2.87e-07$), 24h LBF ($p=3.33e-05$), 4h HBF ($p=9.62e-05$) and 24h HBF ($p=0.000211$).

This same pattern was also observed in the L-leucine metabolite. L-leucine and L-isoleucine are metabolites involved in the valine, leucine and isoleucine degradation pathways, as well as the aminoacyl-tRNA biosynthesis pathway amongst others.

L-tryptophan was spent to a higher degree in the FCS compared to RPMI at 24h in both strains, however, this was found to be significant in the HBF strain only ($p=3.07e-05$). Creatinine was spent in the RPMI media at 4h and was secreted in the FCS media in both strains and found to be significant in both strains at 4h (LBF $p=3.60e-09$, HBF $p=1.08e-10$). Tryptophan is found in the tryptophan metabolic pathway and in the phenylamine, tyrosine and tryptophan metabolic pathways in *C. albicans*. It is also found in the amino-acyl-tRNA biosynthesis pathway in addition to many other pathways. Amino acid data is shown as a bar chart with significance shown in Figure 4.8. Creatinine is additionally found as an intermediate of the arginine and proline metabolic pathways in *C. albicans*. L-arginine was found to be spent in all sample cultures however this was more so in the case of FCS at 24h. L-arginine was significantly more spent or utilised in the FCS media compared to the RPMI at 24h (LBF $p=0.000956$, HBF $p=4.24e-08$). Both L-alanine (LBF $p=1.88e-08$, HBF $p=7.21e-09$) and beta-alanine were found to be more highly secreted in the media at 4h in the FCS media however there was found to be no significant differences at 24h. A similar pattern was observed in 5-oxyproline with significantly higher level of metabolite being found to be secreted in the FCS media at 4h (LBF $p=4.37e-13$, HBF $p=4.37e-13$) and 24h (LBF $p=4.37e-13$, HBF $p=4.37e-13$) compared to RPMI. L-arginine is involved in the arginine and proline metabolic pathway as well as the aminoacyl-tRNA biosynthesis pathway. L-alanine and beta alanine found in the alanine, aspartate and glutamate metabolic pathways in addition to many others including amino-acyl-tRNA biosynthesis and cysteine and methionine metabolism. O-acetyl-carnitine was absent in the media controls for RPMI and was not present in any of the RPMI samples. It was found to be secreted to a small degree at 4h in FCS in both strains and then spent to a significant level at 24h (LBF $p=4.13e-05$, HBF $p=2.90e-07$). It is involved in the degradation of fatty acids and fatty acid metabolism, and specifically facilitates the movement of acetyl-CoA.

Other amino acid metabolism is shown as bar charts with significance between RPMI and Serum indicated in Figure 4.9.

4.4.6 Identified Nucleotide metabolism

Of the metabolites identified under the super pathway of nucleotide metabolism changes in level of metabolite were observed compared to the media control in the FCS media however no secretion or uptake was observed in the RPMI only (Figure 4.9). Hypoxanthine, a component of purine metabolism in *C. albicans*, was found to be significantly spent in the FCS media at 24h (LBF $p=8.05e-10$, HBF $p=3.05e-06$), whereas there was no change in level compared to media control in the RPMI in either strain or time point. This metabolite was also found to be completely absent in the RPMI media but was found in the FCS media control. Cytidine was similarly absent in the RPMI controls whereas it was present in the FCS media control. Cytidine a component of pyrimidine metabolism was secreted to a significant level compared to media control and compared to RPMI at both 4h and 24h and in both strains ($p=4.37e-13$). The same trend and level of significance was found of allantoin in FCS compared to RPMI. Allantoin was also found to be completely absent in the media control. Allantoin is a by-product of nucleic acid and ureic acid decomposition involved in the KEGG purine metabolism pathway.

4.4.7 Carbohydrates

Only 2 compounds classified as part of the carbohydrate super pathway designation within the KEGG database were found to be significant these were citrate and cis-aconitate when comparing supernatant of *C. albicans* grown in FCS to RPMI (Figure 4.10). Citrate was highly secreted in the 4h FCS media by *C. albicans* whereas it was spent in the RPMI. However, the relative concentrations of citrate in the media controls were relatively low to begin with. This difference was found to be significant in both strains (LBF $p=0.00242$, HBF $p=4.38e-05$) However, there were no differences found between media at 24h. cis-aconitate was also found to be much more excreted by *C. albicans* at 4h in the FCS compared to the RPMI (LBF $p=5.12e-10$, HBF $p=6.50e-09$) this was true for both strains however this was not found to be true at 24h.

4.4.8 Identified lipid and fatty acids

Of the identified metabolites, 1 fatty acid and 3 lipid super-pathway related metabolites were found to be differentially excreted or spent within our media between our two strains and our two growth medias at 4 or 24h (Figure 4.10). L-carnitine a facilitator of fatty acid metabolism and specifically the transport of acyl-CoA following β -oxidation. Its level of regulation in FCS being almost identical to O-acylcarnitine. It was found to be significantly different than RPMI in both strains at 24h (LBF $p=0.00375$, HBF $p=6.88e-05$) with it being spent in the FCS compared to the RPMI. It was however secreted to a low level in the FCS and in the RPMI at both time points. There was no change from the media control in the HBF grown in the RPMI media. Taurine is found within the taurine and hypo taurine metabolic pathways in *C. albicans*. It was found to be excreted significantly more in FCS ($p=4.37e-13$) compared to RPMI where it was absent. The level of expression was visibly greater at 4h than 24h (Figure 4.10). sn-glycero-3-phosphocholine a metabolite of the glycerophospholipid metabolic pathway was found to be significant when comparing FCS to RPMI at all time points and in both strains. It is excreted significantly more after 4h (LBF $p=0.000189$, HBF $p=0.00981$) of incubation in RPMI and was spent to significant degree at 24h (LBF $p=0.0139$, HBF $p=0.0115$) in contrast to secreted in FCS. Choline is a precursor in phospholipid biosynthesis in *C. albicans*. Choline was found to be spent in the media by *C. albicans* in all sample groups compared to the media control, however, this was to a significantly greater degree in the RPMI media at 24h (LBF $p=4.37e-13$, HBF $p=4.37e-13$).

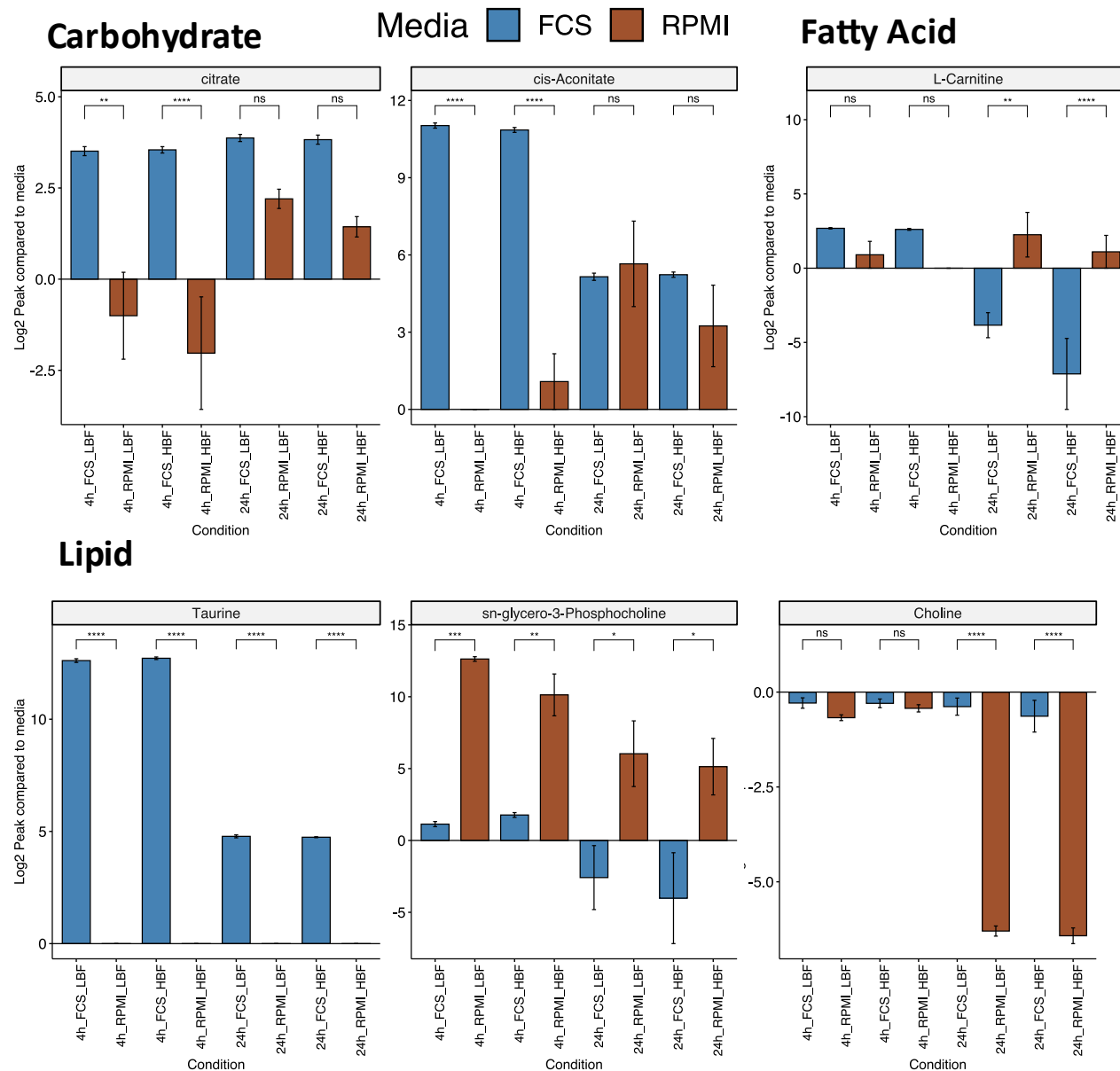


Figure 4.10 Identified metabolites between RPMI and Serum in *C. albicans* isolates. Bar charts depicting the significantly different carbohydrate, fatty acid and lipid related metabolites selected by Analysis of Variance (ANOVA) followed by adjusting the p-value using Benjamini-Hochberg FDR. Identification of significant changes was performed using the Tukey HSD. Values are log2 peak area compared to media control with positive values indicating secretion and negative values indicating that the metabolite was spent compared to media control. For each plot significance is indicated according by the Tukey derived p-value where there is a significant difference between the log2 relative area in RPMI vs Serum in either HBF or LBF and at either 4 or 24h. Significance is indicated between the four comparisons as ns, $p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$ or $p < 0.0001^{****}$.

4.4.9 Pathway activity profiling of *C. albicans* metabolic footprint

Pathway activity profiling was applied to our datasets. The activity of *C. albicans* pathways was predicted from their extracellular abundance data. This data comprised all peaks, both annotated and identified from our targeted/untargeted metabolomics analysis. Pathway activity profiling is dependent upon KEGG identifiers due to the database being used was constructed from KEGG API. Pathways were filtered to include only those in the *C. albicans* KEGG database. This avoids misleading results caused by identification of enriched pathways that are not likely or implausible. Peak IDs were matched to KEGG identifiers and a matrix of samples/IDS against log₂ media normalised peak abundances was utilised for PAPI analysis. Using the values relative to media is recommended for metabolic footprint or extracellular metabolomics by the authors of this methodology. It improves interpretability especially considering that our comparisons are between the metabolomic profiles of two separate medias. It allows for activity to be identified as either secretion or utilisation (spent) of metabolites within a pathway.

An activity score with a positive value would indicate that this pathway activity is due to secreted metabolites to the media, and a negative score corresponds to a pathway activity relating to uptake of metabolites from the media or spent metabolites. Pathways that were different between any of sample groups are shown in as a grid plot (Figure 4.11). These pathways were filtered on a database of known pathways that can possibly occur in *C. albicans* KEGG specific pathways. This was a custom filter added to the analysis to further improve interpretability. As shown with the PiMP based enrichment metabolites are assigned to all possible KEGG pathways as opposed to those that are possible for that organism. Pathways that are theoretically possible from the KEGG database were further screened to identify those that were specifically different in FCS compared to RPMI. ANOVA was performed to identify significantly activate pathways within our entire sample group. Significantly changing pathways with a corrected p-value of ≤ 0.05 according to ANOVA are shown Figure 4.11. These pathways were then further filtered to identify those that specifically changed between RPMI and Serum. Pathways were found to both be significantly different between our sample groups and significantly different between our two media groups. This was ascertained with the Tukey's HSD ≤ 0.05 (Figure 4.12-4.13).

Bar plots were drawn to illustrate the differences in pathway activation and their significance for the top 16 significantly changing pathways that were significant between FCS and RPMI in either HBF or LBF (Figure 4.12-4.13). The analysis methodology for first identifying significant terms by ANOVA and then identifying specifically those that differ between FCS and RPMI by Tukey HSD was similar to the identification of significantly changing targeted metabolites as outlined earlier in section 4.4.4.

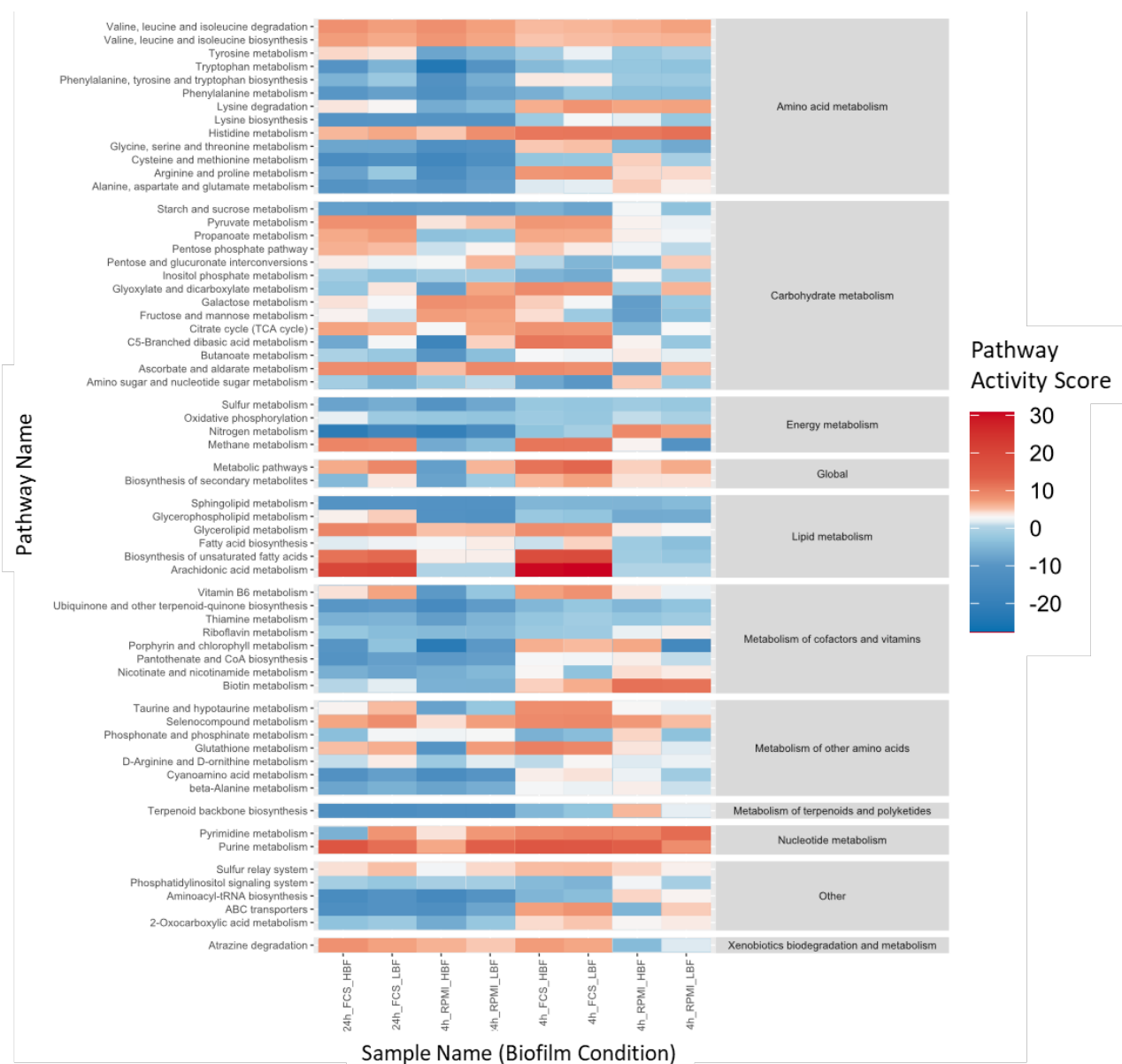


Figure 4.11 . **Grid plot of the level of pathway activation according to Pathway Activity Profiling (PAPi).** Summary of activity by relative square root transformed activity score. The grid plot contains pathways found to be significant between any of our sample groups by ANOVA followed by Benjamini-Hochberg FDR. Positive values indicate that the pathways activated where through secretion of metabolites and negative values indicate that pathways were activated through acquisition or utilisation of metabolites. The grid plot gives a summary overview of the differentially activated pathways in each of our sample groups.

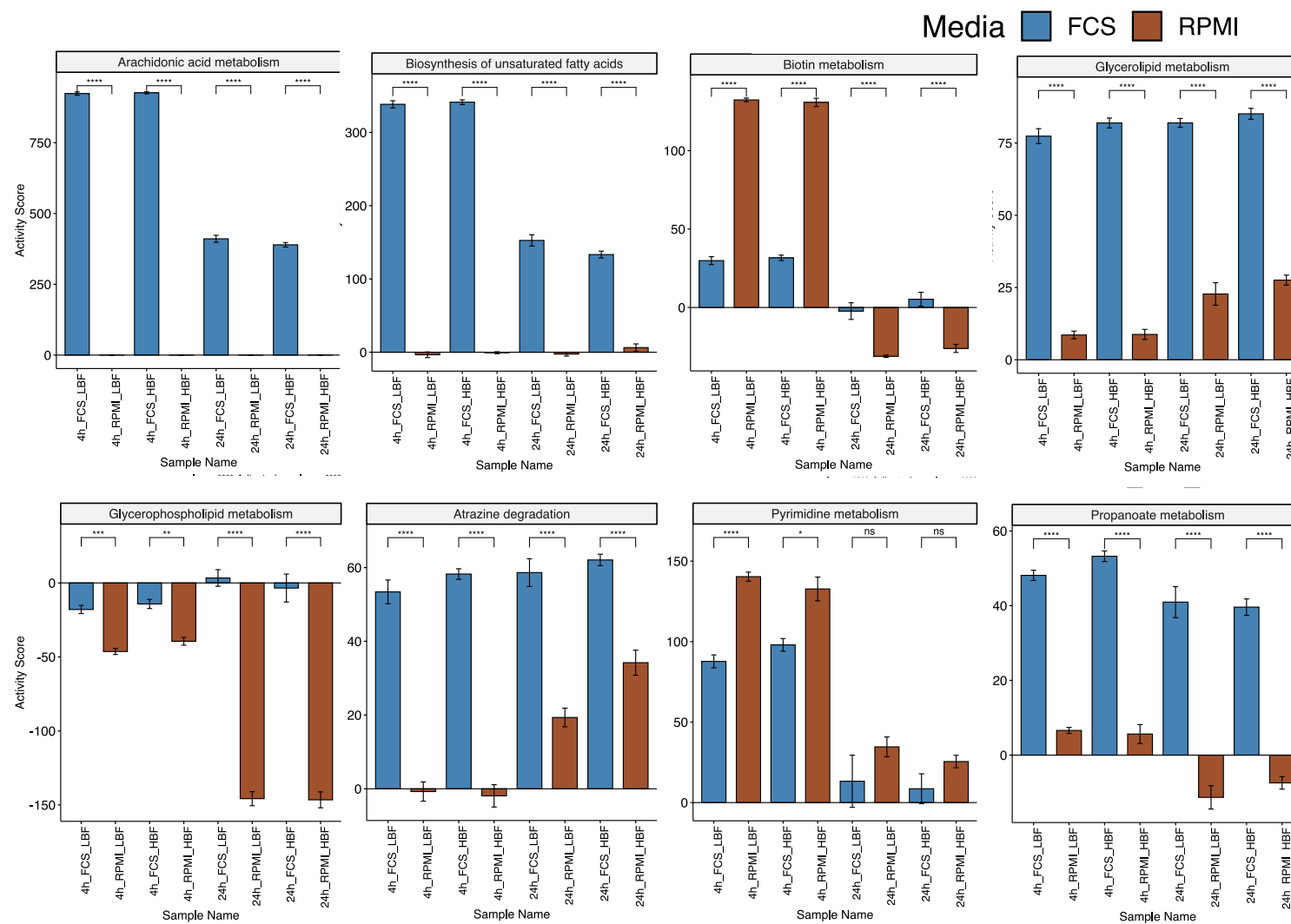


Figure 4.12 **Top 8 activated pathways according to PAPI.** Bar charts depicting the first 8 top significantly activated pathways according to PAPI followed selected by Analysis of Variance (ANOVA) followed by adjusting the p-value using Benjamini-Hochberg FDR. Identification of significant changes was performed using the Tukey HSD. Values are the activity score assigned by PAPI metabolomic pathway analysis algorithm against the KEGG database. Activity scores are either +ve indicating metabolites within that pathway were secreted to the extracellular media or -ve indicating they were spent from the extracellular media. For each plot significance is indicated according by the Tukey derived p-value where there is a significant difference between the activity score in RPMI vs Serum in either HBF or LBF and at either 4 or 24h. Significance is indicated between the four comparisons as ns, $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ *** or $p < 0.0001$ ****.

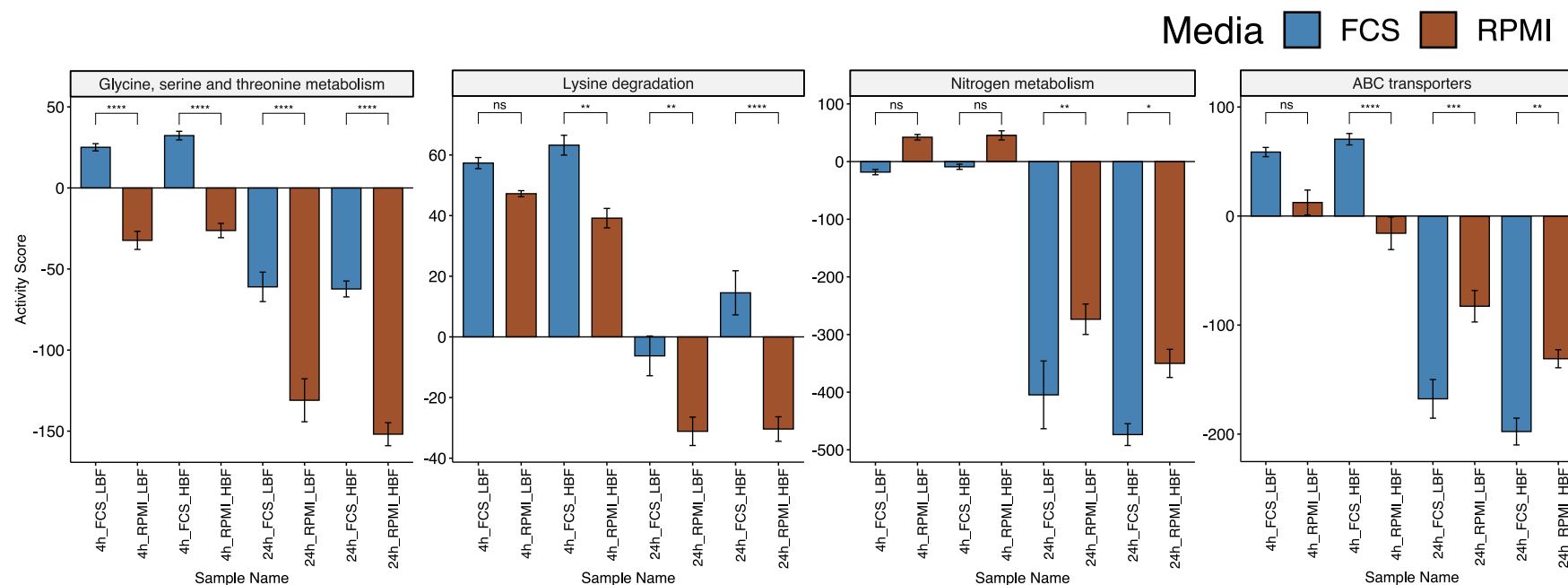


Figure 4.13 Second top 4 activated pathways according to PAPI. Bar charts depicting the second top 4 significantly activated pathways according to PAPI followed selected by Analysis of Variance (ANOVA) followed by adjusting the p-value using Benjamini-Hochberg FDR. Identification of significant changes was performed using the Tukey HSD. Values are the activity score assigned by PAPI metabolomic pathway analysis algorithm against the KEGG database. Activity scores are either +ve indicating metabolites within that pathway were secreted to the extracellular media or -ve indicating they were spent from the extracellular media. For each plot significance is indicated according by the Tukey derived p-value where there is a significant difference between the activity score in RPMI vs Serum in either HBF or LBF and at either 4 or 24h. Significance is indicated between the four comparisons as ns, $p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$ or $p < 0.0001^{****}$.

Arachidonic metabolism was the most significantly regulated pathway due to a high abundance of metabolite secreted to the media by our isolates grown in the presence of FCS, and no activity of the metabolites related to this pathway in our isolates grown in the presence of RPMI only. It was found to be significant at both time points in both strains ($p=4.37e-13$). There is a visibly high activity at 4h with a reduction in activity at 24h in FCS compared to RPMI (Figure 4.12). This would seem to indicate a production and then utilisation of the metabolite involved in this pathway. However, arachidonic acid is not produced by *C. albicans* so is likely cleaving arachidonic acid from phospholipids found within the FCS. This may explain the lack of arachidonic in the RPMI controls. This is confirmed by looking at the relative peak abundance data for the arachidonic acid identified peak. There is a high abundance of this metabolite in our FCS supplemented *C. albicans* strain media, but no abundance in our RPMI only media *C. albicans* strain media.

Biosynthesis of unsaturated fatty acids had a much higher positive activity score in both of our isolates grown in FCS which was significant at both 4 and 24h time points ($p=4.37e-13$). More pathway intermediates were present in the media than in the RPMI. This was true for both time points with slightly more activation of this pathway at 4h in the FCS. This difference was significant at both 4 and 24h and in both strains between RPMI and FCS. Biotin metabolism was shown to have a similar trend in all conditions with the metabolites secreted at 4h (LBF $p=4.37e-13$, HBF $p=4.37e-13$) and then either less secreted in FCS or spent in the case of RPMI at 24h (LBF $p=3.97e-06$, HBF $p=1.11e-08$). The difference was significant at all time points and both strains between FCS. Fatty acid biosynthesis is an integral process involved in the process of producing fatty acids for energy storage, constituents of the cell and protein modification. Biotin is required in *C. albicans* for growth and fatty acid metabolism.

Glycerolipid metabolism had a higher positive activity score in all FCS samples compared to RPMI which was deemed to be significant at all time points ($p=4.37e-13$). A positive activity score (AS) is indicative of metabolites from this pathway being secreted into the media by *C. albicans*. The same trend was apparent in the pyruvate metabolism, which was shown to have a higher AS, and therefore higher abundance of pathway metabolites in the extracellular media in FCS compared to RPMI. This was found to be true in both strains and in both time

points. However, this difference was less pronounced at 24h with the AS being higher in RPMI at 24h compared to 4h. Pyruvate metabolism is pivotal metabolic pathway in yeast. Pyruvate is a ketone that is a junction in metabolism produced from glycolysis and can be converted to carbohydrates or utilised in the production of fatty acids. Glycerophospholipid was shown to have a difference in uptake activity in both our strains in RPMI compared to FCS in both strains at 4h (LBF $p=0.000973$, HBF $p=0.00302$) and 24h (LBF $p=4.37e-13$, HBF $p=4.37e-13$). Propanoate pathway activity showed a similar pattern in both strains at all time points however was significantly different between FCS and RPMI.

Atrazine degradation pathway activity was significantly more positive in the FCS media at 4h (LBF $p=4.39e-13$, HBF $p=4.37e-13$) and 24h (LBF $p=3.44e-06$, HBF $p=4.28e-07$) in both strains. This is indicative of a higher level of pathway metabolite intermediates present in the media point. This difference was found to be significant in both strains in FCS compared to RPMI. A very similar pattern of activity and significance was observed for pyruvate at 4h and 24h with $p=4.37e-13$.

Glycerine serine and threonine metabolism showed fluctuation in activity between our groups. With it being significantly more excreted at 4h (LBF $p=9.76e-06$, HBF $p=2.88e-06$) and significantly less spent at 24h (LBF $p=9.07e-11$, HBF $p=5.51e-10$) in the FCS media.

Lysine degradation was significantly activated at 4h (LBF $p=0.656$, HBF $p=0.00162$) in HBF but not LBF. Both medias showed positive activity however this was more positive in the HBF in FCS. At 24h the pathway was found to be negatively activated with the pathway metabolites being spent except for in the case of HBF grown in FCS which had a positive activity score. The level of activity was significantly different between RPMI and FCS in both strains at 24h (LBF $p=0.00363$, HBF $p=4.30e-08$).

In general, as indicated by our PCA analysis the level of pathway activity and metabolite regulation was found to be consistent between our two strains. Some differences were observed such as alanine, aspartate and glutamate metabolism specifically between the strains grown in FCS and lysine biosynthesis. Fluctuations in many important metabolic pathways were observed to be distinguished in the *C. albicans* in the FCS supplemented media

compared to RPMI alone. The most significant feature was the arachidonic acid, but this may be due to identification of high levels of metabolite in the FCS and complete absence in the RPMI despite its low representation of metabolites. However relatively high levels of arachidonic acid were observed in the peak abundance data in the samples grown in FCS (Figure 4.13). Arachidonic acid pathway enrichment is however only enriched by arachidonic acid. As noted earlier arachidonic acid is not produced by *Candida* but could be cleaved from phospholipids by phospholipases. *Candida* utilises host derived phospholipids, which may be present in serum, to derive arachidonic acid. This in turn can be used as a carbon source and production of eicosanoids (Mishra et al., 2014, Ells et al., 2008).

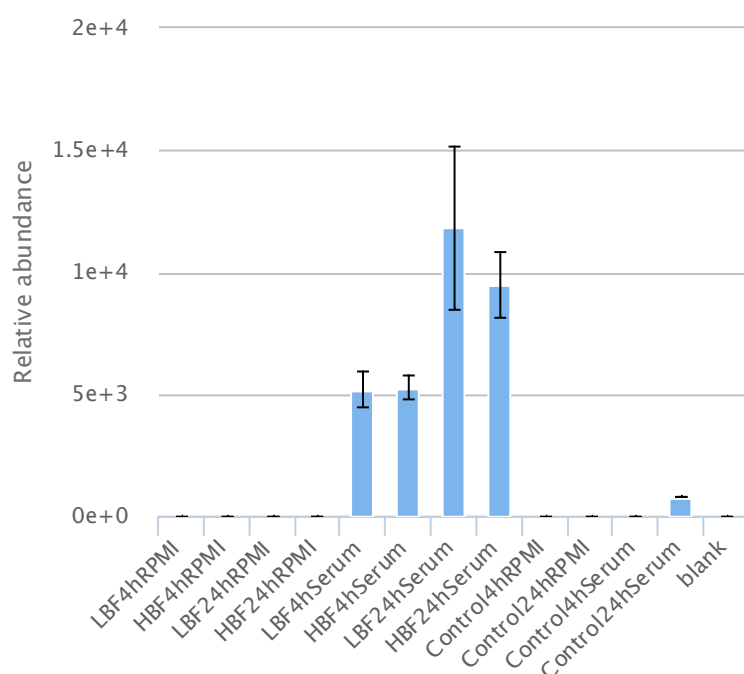


Figure 4.14 **Bar chart comparing the relative abundance of arachidonic acid.** Relative abundance of Arachidonic acid as discerned by the PiMP processing pipeline. Metabolites are representative of levels detected in the spent media of high and low biofilm *C. albicans* isolates in the grown in presence or absence of FCS and at 4h and 24h incubation times. Relative abundance with standard deviation is plotted for each sample group noted as isolate type, followed by the time the biofilm was grown for followed by the media it was grown in.

4.5 Discussion

This research aimed to apply the widely targeted/untargeted pHILIC/LCMS to interrogate the metabolomic footprint of *C. albicans* clinical isolates in the presence and absence of foetal calf/bovine serum (FCS). The clinical strains were categorised as either HBF or LBF, as previously described in Scottish Candidemia studies performed by Rajendran et al (2012) at the Glasgow Dental Hospital (Rajendran et al., 2016d). We aimed to compare the metabolic footprint of these high and low biofilm forming isolates in response to serum which is a known agitator of hyphal switching and morphogenesis in *C. albicans* (Villa et al., 2020). Hyphal morphogenesis is a feature of *C. albicans* heavily associated with biofilm phenotype. Additionally, we wished to contrast the difference in the impact of FCS on the LBF compared to HBF. From *in vitro* observations, highlighted previously in chapter 3, the biofilm inducing effects are more pronounced in the biofilm deficient or low biofilm forming strains.

We use the term metabolic footprinting to distinguish it from metabolic fingerprinting, which is the measure of intracellular metabolite levels. Metabolic footprinting within the research performed here is the measure of levels in the supernatant or spent media of the organism (Kell et al., 2005). Fluxes in the metabolic activity of the cell is reflected within the spent media through extracellular changes in secreted and spent metabolites. Metabolomic footprinting was chosen due to the nature of the work looking at the metabolism of components of the two medias. The initial aimed wished to look at the specific components of the serum media which were utilised by the *C. albicans* biofilms. Metabolic footprinting offers a simple procedure for extraction of metabolites and inference of the phenotype of microorganisms. However, it is less informative in comparison to some other metabolomics methodologies. Metabolomic footprinting provides a proxy for interpretation of intracellular activity, however, it is only a proximally, and the method could be improved with the integration of the intracellular metabolome or the fingerprint and even further improved with the use of metabolic flux by stable isotope labelling (Han et al., 2019b). Intracellular metabolomics is more commonplace and in combination with footprinting may have served to fill in gaps. This combined intracellular/extracellular profiling approach has been performed previously in 2016 by Weidt et al in *Candida* biofilms (Weidt et al., 2016). Analytes that are not externalised during growth would not be detectable, and therefore a combination approach would offer

greater insight and may explain incompleteness of metabolic pathways. Metabolic flux analysis would also offer improved resolution through characterisation of reaction rates of enzymes within the biofilms (Sauer, 2006). Commonly ^{13}C labelled carbon atoms are utilised to detect rates and fluxes of networks through tracking the abundance of these carbons through the metabolic pathways. This method would allow you to track rates of metabolic networks between the two strains or between two growth medias. Metabolic flux such as that using ^{13}C labelled carbon has been demonstrated within *C. albicans* to identify metabolic activity specific to hyphal formation (Han et al., 2019b).

The footprint often referred to as the exometabolomic does provide a high level of information of microbial activities specifically under different culture conditions in comparison to other more limited fluorometric assays. Within our research it has proven sufficient to distinguish differences in the metabolome under two specific culture conditions (Sue et al., 2011). We used 3 different high and 3 different LBF strains and 3 technical replicates which showed a great deal of overlap and reproducibility in their metabolic profiles. This was confirmed by our PCA analysis which showed that our metabolic footprints were reproducible for our 3 strains. Interestingly there was also the two phenotypically different strains were similar in their metabolic profiles. The samples were distinguishable by their time point of collection and by the media they were grown in. Initial observations would seem to imply that exposure to serum alters the metabolome of *C. albicans*, however, this is consistent from strain to strain. Despite failure of a few technical replicates within the mass spectrometry run we retained enough samples adequate for drawing conclusions from our data. Fragmentation data would have likely improved the interpretability of our untargeted analysis. Fragmentation data or ms/ms spectrum data would have allowed resolution of some of analytes which are ambiguous within our untargeted data (Pitt, 2009).

When comparing LBF to HBF in comparison with the previous studies, we have performed within the group. L-phenylalanine, L-glutamate, L-tryptophan were identified as key differentiators of HBF compared to LBF. We previously identified that AAT1 was at the centre of network of genes linked to amino acid metabolism (Rajendran et al., 2016a). Specifically, phenylamine, tyrosine and tryptophan metabolism and alanine, aspartate, and glutamate metabolism. It is known that amino acid uptake such as L-phenylamine has a functional role

in germ tube formation (Kaur and Mishra, 1991). With uptake of L-phenylamine being much greater in full hyphae compared to non-hyphal cells in these early studies. Amino acid uptake is related to the pH regulation in *C. albicans*, pH in turn is a regulator of dimorphism in *C. albicans*. L-glutamate is a preferred nitrogen source for yeast followed by less preferable amino acids such as L-tryptophan, however both were found to be distinguishing features between LBF and HBF (Kraidlova et al., 2016). These amino acids were in higher concentration in the LBF indicating that they are not being utilised to the same degree as HBF.

This data perhaps helps to inform and confirm our earlier observations in 2016 that amino acid metabolic processes are important in biofilm formation (Rajendran et al., 2016a). Additionally, these processes are seemingly activated in HBF to a greater extent than LBF. PLS was able to identify these key differentiators between the high and low biofilm formers with many of the differentiating features between HBF and LBF being involved in amino acid and carbon metabolism. Identification of changes between HBF and LBF was based upon targeted metabolomics which could influence the analysis towards amino acid metabolism due to the standards used. From the metabolic footprint data, it was possible to determine that there are differences in the metabolism between HBF and LBF. However, it is difficult to draw conclusions about how these metabolites are being utilised by the cells. Fingerprinting and/or confirmatory experimentation would be needed to discern a greater understanding of amino acid metabolism in these phenotypically different isolates.

Targeted analysis based upon known standard with accurate mass and retention time successfully identified 49 compounds within our samples. Additionally, we identified 691 putatively assigned metabolites using only mass and no standards which were then annotated with corresponding KEGG IDs. We performed a direct statistical comparison between our groups on the identified metabolites, this was possible due to our increased confidence in the identity of these metabolites. Annotated metabolites due to their more spurious nature were considered and visualised as contributors to entire metabolic pathways in enrichment type analysis. PAPI analysis was employed as this methodology is well suited to metabolic footprinting due of its ability to visualise both secreted and spent pathways activities. Pathway analysis software's such as PAPI are dependent upon *a priori* knowledge maintained in databases such as the KEGG databases. R packages such as PAPI and other GUI based software

can access these databases through their APIs. By subtracting the observed analyte values from the media before undertaking PAPI allowing for activity scores to be calculated as positive or negative relative to the media. This allows not only for a determination of activity relevant to other samples but also to see the overall direction of the activity, whether that be an increase in secretion or uptake of metabolites. This feature of PAPI for metabolic footprinting added an extra layer of interpretability to the pathway findings.

Our targeted analysis using the identified metabolites revealed numerous changes within our list of standard compounds between the FCS supplemented *C. albicans*. Significant changes in amino acids were observed, with the majority of significant metabolites belonging to this family, which in part is due to the larger representation of amino acids within the standards used. Those spent over time by *C. albicans* in FCS included betaine, L-glutamine, L-asparagine, L-isoleucine and L-leucine. Those spent more in the FCS un-supplemented included L-tryptophan and creatinine.

Amino acid sensing and uptake in *C. albicans* is a known inducer of hyphal morphogenesis. The amino acid sensor *ssy1p* (*csy1p*) is nutritional response sensor that activates the amino acid permease AAP genes (Klasson et al., 1999). AAP genes are responsible for transcription of amino acid permease, a superfamily of proteins responsible for transport of amino across the plasma membrane (Martínez and Ljungdahl, 2005). The regulation of these genes has been previously demonstrated to be dependent upon sensing through the *ssy1p*. The SPS plasma membrane sensor, comprised of the subunits *Ssy1*, *Pr32* and *Ssy5*, activated in fungi upon sensing of extracellular amino acids results in proteolytic cleavage of the transcription *Stp2* inducing genes responsible for amino acid metabolism (Böttcher et al., 2020). This mechanism in *C. albicans* is associated with virulence, triggering of yeast to hyphae morphogenesis and biofilm formation (Miramón and Lorenz, 2016, Ramachandra et al., 2014). Several amino acids were differentially changing in FCS in both our low and high biofilm formers including L-arginine, L-alanine, beta-alanine, s-oxyproline and L-carnitine. These amino acids are related to several KEGG pathways within *C. albicans*. Amino acids are thought to trigger signalling pathways in *C. albicans* either directly or indirectly such as Arginine in the activation of SPS plasma sensor and or indirectly through their biosynthesis leading to alkalinisation (Garbe and Vylkova, 2019). Amino acid pathways including cyanamino acid

metabolism, glycine, serine and threonine, and others are activated differentially between FCS compared to RPMI.

From the identified metabolites related to nucleotide metabolism many failed to be identified in any of the RPMI samples. Serum contained higher levels of hypoxanthine, which were spent at 24h. Citidine was found to be excreted in the FCS media. Hypoxanthine is a purine base related compound, transported by *C. albicans* by the Fcy21p adenine-guanine-hypoxanthine-cytosine permease (Pantazopoulou and Dailianas, 2007). *C. albicans* can utilise purines as a nitrogen source. The Fcy21p expression is thought to be linked to nitrogen starvation as demonstrated by Goudela et al (Goudela and Tsilivi, 2006). Citrate from the super pathway of carbohydrate was secreted more by the *C. albicans* grown in serum at 4h at which point the *C. albicans* strains in RPMI had spent this metabolite. Higher levels of the metabolites citrate and aconitate indicate higher level of TCA at the 4h time point in *C. albicans*. However, citrate to cis-aconitate is a reversible reaction so directionality is not deducible from these identified metabolites. The TCA cycle was shown to be more active at the earlier biofilm time points in the FCS than RPMI from the PAPI analysis. Mature biofilm formation is linked to the essential activation of both the amino acid and TCA metabolic pathways (Tao et al., 2017). Pyruvate was activated to a higher degree in *C. albicans* metabolism through the putative analyte secretion lactate and malate.

Biosynthesis of unsaturated fatty acids and the related propanoate metabolic pathway were highly activated in FCS compared to RPMI putatively assigned compounds arachidonic acid, docosahexadonic acid, oleic acid and stearic acid. Arachidonic acid and oleic acid play an important role in the synthesis of prostaglandins. Arachidonic acid is a precursor for the production of Prostaglandin E2 (PGE2) (Mishra et al., 2014). Production of prostaglandins by *C. albicans* has been shown to modulate hyphal formation and control the host inflammatory response (Noverr and Huffnagle, 2004). Arachidonic acid was visualised noted as it was found to be a very serum specific metabolite, having no activity within the RPMI only or serum deficient cultures. Future work will need to consider *in vitro* experimentation to determine the quantity and related enzymatic activity within a larger population of strains to determine the relationship between biofilm formation and hyphal morphogenesis in *C. albicans*. Arachidonic acid was noted as it was found to be a very serum specific metabolite having no

activity within the RPMI only or serum deficient cultures. Arachidonic is typically reported as being an exogenous metabolite utilised by fungal species in the production of prostaglandins (Mishra et al., 2014, Liu et al., 2016). As we have indicated *C. albicans* does not produce arachidonic acid and is likely deriving this from phospholipids within the FCS. These phospholipid biproducts could be the nutritional and biosynthesis requirements for LBF biofilm development. This is interesting as these are host derived metabolites.

Glycerolipid intermediates were secreted in the FCS and the glycerophospholipids were more spent in the RPMI. These pathways were putatively annotated by their intermediate's glycerol, Glycerone, D-glycerate and D-glyceraldehyde. Glycerophospholipids are synthesised from products of fatty acid synthesis for production membrane glycerophospholipids in yeast (Klug and Daum, 2014).

Within this chapter we have identified key metabolites that differentiated between HBF and LBF which confers some of the findings of our previous work. We have also identified interesting links in amino acid, fatty acid metabolic and prostaglandin producing pathways. Many of which are previously linked to hyphal and biofilm formation within the literature. Again, it is interesting to note that our LBF strains mirror the HBF strains in their metabolism under the influence of FCS. Many similarities exist between the two strains in response to serum, however, there are instances in which LBF vary in their metabolic profile in comparison to HBF.

In conclusion, metabolic foot printing identified several pathways and identified metabolites changing in serum and RPMI. This goes to show the metabolic flexibility and plasticity undertaken by *C. albicans* in response to nutrient and environmental stress. More focused approaches to see the activation of specific pathways as discussed using metabolic flux analysis and removal of ambiguity of ms/ms would improve this study design. Within the current study putatively assigned metabolites are responsible for the enrichment of numerous pathways. However unlike in transcriptome-based studies, which also use enrichment, these metabolites are also identified from the same peaks as other metabolites. Pathway analysis, however, serves our interpretation well as we can hypothesise pathways that are being reprogrammed in FCS by our strains. What we observed from our

transcriptome data in chapter 4 was that there were numerous differences in up and downregulated differences in pathways in \pm FCS. From our multivariate analysis, individual analysis of identified metabolites and through analysis of differentially activated pathways we observed that there were fewer differences between HBF and LBF in response to serum in the metabolomics data. This may be due to the methodology of metabolic footprinting as we have discussed in that we do not see what is happening within the organism. Therefore, it would be interesting to integrate the pathways and metabolic and transcriptomic profiles. A few methodologies have been developed to approach this topic as touched upon within Chapter 1 and will be explored in greater detail in the next chapter.

4.6 Highlights

- Identification of amino acid metabolism and related pathways in HBF compared LBF.
- Distinct metabolomic profile of *C. albicans* isolates in FCS compared to RPMI.
- Metabolic footprint is modified by activation of Amino acid, Lipid and Carbohydrate pathways.
- In contrast to transcriptome fewer distinguishing features between HBF and LBF in the presence of FCS.

5 Omics Integration and Discussion

5.1 Introduction

Within the human host the switch from free floating planktonic cells to surface-attached and aggregated biofilm communities pose a particular clinical concern. Advantages offered to biofilms through this phenotype include increased tolerance to antimicrobial therapies and difficulties in clinical management. Like many bacterial pathogens it has been observed that the biofilm phenotype is observed within fungal species like *C. albicans*. We have shown through the literature that virulence and resistance has been a key feature underpinning the direction of *C. albicans* biofilm research, with a growing interest in interkingdom interactions. Work has been undertaken to test this phenotype *in vitro* and to interrogate the underlying mechanisms driving biofilm formation. As with many areas of academia, as research progresses in one area then this opens avenues of investigation in others. Indeed, it has been observed within microbial populations, including *C. albicans*, that not all strains and populations behave the same. We have observed within clinical cohorts that biological heterogeneity with respect to biofilm formation and resistance is a prominent feature in patients (Rajendran et al., 2016d, Rajendran et al., 2016a). These higher levels of antifungal resilience and biofilm forming ability is associated with increased mortality through clinical observations (Rajendran et al., 2016d). This thesis sought to explore this heterogeneity using a range of computational approaches to better understand the direction of research within the *C. albicans* biofilm community, which species it interacts with in biofilm infections, and the mechanisms underpinning biofilm formation.

5.2 Understanding the biogeography of biofilms

The field of microbiology continues to progress rapidly in no small part due to the rate of growth and cost effectiveness of omics technologies. Sequencing, mass spectrometry and NMR based technologies have propelled our hypothesis driving methodologies through providing holistic and high throughput profiling of molecular changes within microbial populations. NGS, most utilised in genomics, amplicon, shotgun sequencing transcriptomics within microbiology, has gone from being from expensive and niche to inexpensive and commonplace within the field. The availability of genome assemblies of bacterial and fungal pathogens has grown with the availability, and although many microbial genomes exist in

different stages of assembly, long read assembly methods through PacBio technologies and the Nanopore sequencers offer possibilities of increasing these assemblies further. 1000s of prokaryote genomes have been sequenced, and this number had grown to around 28,000 in 2014, around 70,000 in 2017, and as of 2020 this number has grown closer to 200,000, with nearly 20,000 of these being considered complete genomes (Tatusova et al., 2014, Parks et al., 2017). The rise in genome databases has become an invaluable tool for amplicon-based and shotgun sequencing microbiome studies, and has allowed for taxonomic identification against the vast database of genomes (Balvočiūtė and Huson, 2017). Fungal amplicon or mycobiome studies, utilising the hypervariable ITS region, are starting to appear more readily in the literature (O'Connell et al., 2020, Richardson et al., 2019). These are more limited than bacterial 16S microbiome studies, though comparable in nature, due to our lesser knowledge base in fungi and the maturity of bioinformatic methods to analyse them (McTaggart et al., 2019). 16S amplicon studies have benefited from the field developing over time and improved analytics (McTaggart et al., 2019). The number of genome assemblies for fungal species in comparison is only 348, within the refseq database, at the time of writing, with only 15 of these described as complete (O'Leary et al., 2016, 2020). Initiatives such as the 1000 fungal genomes project have been progressing this and aim to address the under representation of fungal genomes (<http://1000.fungalgenomes.org>).

It is clear that *C. albicans* does not exist in isolation, and that it often resides within a complex interkingdom biofilm in a number of sites within the human host (Nobile, 2013). Due to numerous advancements within microbiology, it has become apparent that diseases are influenced by communities of organisms. These discoveries have been aided in no small part through NGS and 16S amplicon sequencing. As we have observed anecdotally, and additionally through reviewing the literature within this thesis, interkingdom relationships are important in the context of *C. albicans* biofilms. Clearly, there are key synergistic and antagonistic relationships that influence community structure and the overall biogeography of mucosal surfaces. The use of high throughput technologies on these interkingdom biofilms has already been applied to dual-species models including *S. gordonii*, *S. mutans* and *P. gingivalis* (Sztukowska et al., 2018b, Dutton et al., 2016, He et al., 2017). However, in general these communities are more complex than these two species models. Future work would aim to build upon these models further, with the addition of multiple organisms simultaneously

drive greater complexity to the biofilm models. The possibility to do this with greater frequency as these technologies move from being specialist facilities and services to everyday benchtop equipment, thereby minimising bottlenecks for innovation and advancement. Understanding these interactions perhaps needs to begin with improved profiling of the important interactions within a given disease.

As a part of this work, we identified relationships that exist between *Candida* spp. and bacterial species within oral sites within the host. There was an important limitation in that qPCR was used for total fungal load to compare to our amplicon data. A better indicator of interactions would have been profiling of the fungal mycobiome, though these experiments predated our current technological capacity for ITS sequencing. The mycobiome has perhaps been overlooked, microbiome studies and their supporting databases have increased dramatically whereas the same is not true of the mycobiome research with methodologies for microbiome being far more standardised (Proctor et al., 2019, Cullen et al., 2020, McTaggart et al., 2019, Tiew et al., 2020). Importantly, future whole community studies should also consider the mycobiome as part of their methodology. Fungal mycobiome studies do exist that highlight their importance in disease and health comparisons such as the differently abundant taxa observed in caries (O'Connell et al., 2020). Bacteriome and mycobiome joint studies also have been performed within the oral cavity, however wider adoption would enable the fungal mycobiome studies to improve and have a greater level of standardisation similar the bacterial microbiome (Li et al., 2019, Peters et al., 2017, Persoon et al., 2017b). Nevertheless, as with many fungal diseases there is a reluctance to accept the notion that fungi play a significant role. Despite this, my own ideas that have evolved from the evidence generated within this thesis is that *C. albicans* plays an important 'active participant role' opposed to an 'innocent bystander' in complex communities. This is supported by studies by the Krom group who suggests that *C. albicans* may serve as a 'keystone commensal' in multispecies oral communities (Janus et al., 2016). It seems logical that incorporation of a fungal element provides physical support amongst other attributes to bacterial members of the community. However, to fully appreciate this we must first decipher how *C. albicans* biofilm formation is co-ordinated before we add complexity and start investigating interkingdom interactions at the molecular level.

5.3 Omics approaches to *Candida albicans* biofilms

To support bioinformatics and the high throughput methods for interrogating molecular biology systems, many protein bioinformatic databases have been developed. Sequence databases, structural, chemistry, enzymatic and ontological to name a few have been developed (Chen et al., 2017). These databases have been developed to address more specialised and more specific bioinformatic databases. For example, the Kyoto Encyclopaedia of Genes which contains a per species database of enzymes and related pathways (Kanehisa et al., 2015). Gene Ontology is a database that was developed to maintain a high-level consistent dictionary of terms (ontologies) to describe molecular functions within systems biology (2006). These terms are also maintained across species for the most part unless highly specific to the organism. Many of these databases, for example UniProtKB, PANTHER, pfam and TIGRFAMS and other family databases are utilised in homology-based methods, such as BLAST+ and HMM to annotate genome assemblies (Apweiler et al., 2004). Fungal specific databases include the *Saccharomyces* Genome Database (SGD) and the *Candida* Genome Database (CGD). The CGD is based upon the SGD and maintains a curated database of genomic and protein information for *Candida* spp. Additionally they maintain orthologous information and cross reference to other protein databases such as UniProt (Skrzypek et al., 2016). These bioinformatic databases improve interpretability of sequencing-based experiments and provide biological interpretability of otherwise insurmountable datasets. These databases are integral to amplicon, RNA-Seq, metabolomics, proteomics, and other omics experiments. The more complete our knowledge base is then the more informed our interpretation of omics data can be. GSEA, ORA and other pathway analysis methods all depend upon this solid a priori knowledge. The growth of these resources also parallels the growth of omics techniques. Collectively, these tools have provided the possibility to interrogate large data sets with greater rigour and enabled us to start integrating data.

These databases can be utilised to perform ORA and GSEA, and this API is utilised by many software's that implement either ORA or GSEA that includes metexplore, PiMP and PaintOmics3. It is through these knowledge bases that many of the functional, pathway and network-based models are possible. Human and mice studies have led the way, such as the

large functional databases held by the GO consortium. Both tools and databases for larger variety of species are becoming available and these are being advanced all the time.

The R package clusterProfiler additionally uses KEGG, which is a wrapper that provides both ORA and GSEA through the R function fgsea. GSEA has been performed on the upregulated genes in serum in both LBF and HBF *Candida* isolates. From this we have ascertained overlapping pathways upregulated in both HBF and LBF (Figure 5.1). What we observe is that over representation analysis shows difference between HBF and LBF and the overlap between the two. LBF genes are significantly enriched in more pathways than HBF with more pathways having an adjusted p-value of less 0.05. It is possible to observe that genes involved in a number of pathways that we observed to have differential activity from our metabolomics footprint. Notably, biosynthesis of unsaturated fatty acids and propanoate and fatty acid metabolism. Many of the pathways that we see overrepresented in the LBF grown in serum compared to RPMI did not appear in our PAPI analysis. The top 10 pathways are illustrated for HBF, LBF and the overlap despite many not being significant in HBF. An explanation for this may be that the LBF is regulated or reprogramming its metabolism to a greater extent than the HBF. Interestingly we also observed upregulation of the MAPK signalling pathway in LBF in serum.

5.4 Integration

Single omic experiments are almost “commonplace” in the year 2020 within life sciences and microbiology. However, progress and breakthrough in technology creates hypothesis and further avenues within research. This is common to all of life sciences including microbiology and more specifically biofilm research. Integration within the field of microbial bioinformatics is one of the current challenges. There is interest in integrating omics of every description including the integration of transcriptomic and metabolomic datasets. Transcriptomics and metabolomics have a complicated, not necessarily direct, relationship which makes integration challenging and interpretation convoluted (Cavill et al., 2015).

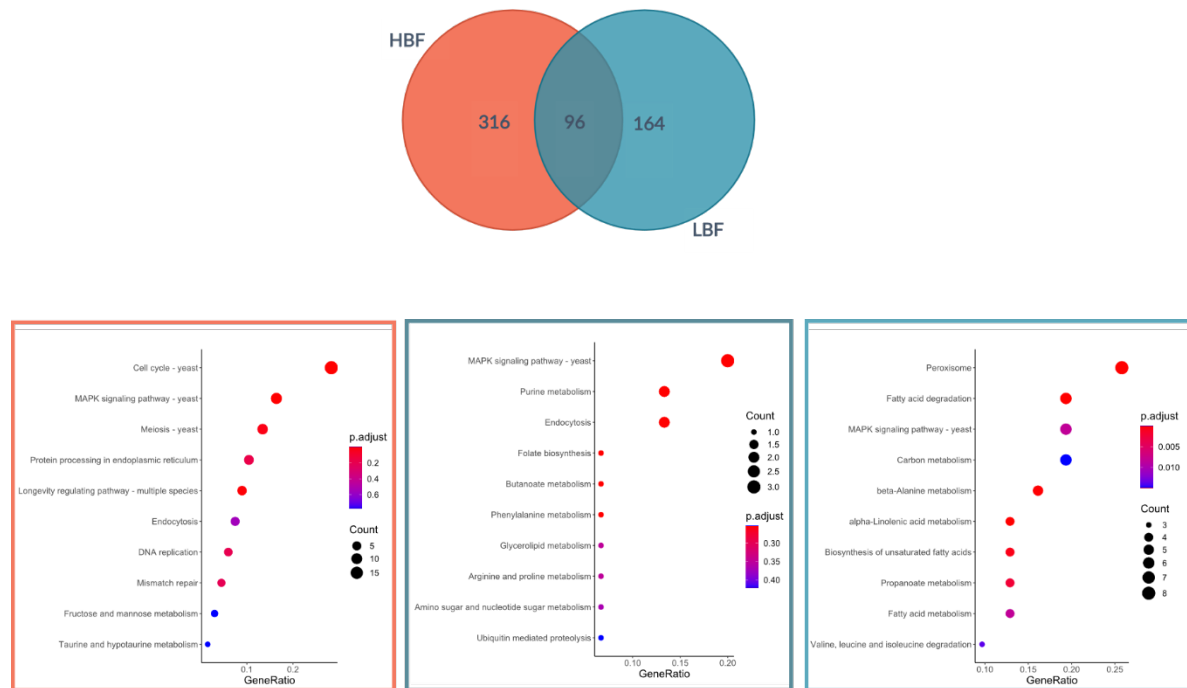


Figure 5.1 Comparative analysis of the enriched KEGG terms serum in both low and high biofilm forming *C. albicans* isolates. Differentially expressed genes that were upregulated in *C. albicans* isolates that were grown in serum compared to RPMI are compared. Those genes that were unique to HBF, LBF or overlapped between the two were submitted to over enrichment analysis against the known KEGG pathways for *C. albicans*. The top ten terms for each analysis are shown although they are not considered significant unless adjusted p-value <0.05. Significance was determined by hypergeometric testing. Dotplots indicate the level of significance of the FDR adjusted p-value (p.adjust) and the size of the dot indicate the number of over enriched genes within the pathway within each analysis.

Methodologies for integrating transcriptomics and metabolomics can be categorised as conceptual, multivariate and pathway. In the coming sections we aim to discuss these methodologies in detail regarding *C. albicans* biofilm research. Using metabolomics and transcriptomics data derived from research in previous chapters we will identify strengths and potentials of these methodologies at interrogating biofilms. Some limitations exist within the data which is important to note. A repeat study methodology was used to derive the samples for the transcriptomics and metabolomics. This introduces a batch effect which due to the nature of the two data types formed is not possible to correct for the inherent batch effects introduced. As the biofilms were grown on separate days from separate medias there are variables that are technical and variability that are unique to the analysis performed in each batch. Some variability can be typically removed within individual technologies, such as gene expression data with which you can normalise data to the same features across batches.

The batch effect is between two different methodologies therefore it is not possible to normalise the metabolomic data in one batch to the gene expression in another batch as they do not share features. This is an important consideration and ultimately limits the power of the data. However, we still wished to utilise the data to assess the current approaches to integrate data sets. The metabolomics data was also derived from a larger sample set, so in certain instances such as multivariate analysis it was considered necessary to reduce the sample set to make it comparable with the transcriptomics. This also introduced a further limitation as we had an incomplete dataset for both the transcriptomic and metabolomic datasets which will be outlined.

5.4.1 Conceptual integration

Conceptual integration is perhaps the simplest method of data integration. In the case of transcriptomics and metabolomics it involves the direct comparison of separately analysed data sets. A recent example investigating the metabolomic reprogramming of *C. albicans* in response to hypoxic stress utilised both metabolomics and transcriptomics (Burgain et al., 2020). The authors identified metabolomic pathways that underwent perturbations in response to hypoxia. Subsequently they visualised differentially expressed metabolites to observe whether these perturbations were also observed in the transcriptome of *C. albicans*. The research does not make any strong claims regarding the effectiveness of this method. However, they highlight transcriptional signatures that overlap with the metabolomic findings. The use of ontologies in the form of metabolomic pathways allows for at least related genes to be integrated. Similarly, a study identifying changes in metabolism of *S. mutans* cultured with *C. albicans* utilises both the transcriptome and metabolome (He et al., 2017). In this instance transcriptomics drives the hypothesis of changes regarding sugar metabolism in *S. mutans*. The usefulness of metabolomics is in its ability to test the extracellular concentrations of metabolites to observe this sugar utilisation. Whether or not this is strictly regarded as integration, it nevertheless has an additive usefulness to utilise both technologies to improve the completeness of the interpretation.

We performed conceptual integration of our metabolomic and transcriptomic data. We did this based upon key pathways that had been identified through PAPI in our metabolomics analysis. A list of pathways with differential activity scores was compiled from analysis in

chapter 4. From our list of metabolic pathways, we then identified transcripts which were differentially changing between RPMI and serum at 4h or 24h and in HBF or LBF. The transcripts which were differentially expressed were filtered and considered significantly changing according to original cut-off of $\geq 1.5 \log_2$ fold change and adjusted p-value ≤ 0.05 . A custom script within R was used to filter by differential expression and then by whether those transcripts were present in our significantly changing pathways from our PAPI analysis as determined by FDR ANOVA $p < 0.05$. Of our 51 pathways that were significantly changing in serum compared to RPMI in either of our strains, 46 of those pathways that contained genes that were significantly changing within those pathways. The core metabolic pathways and their corresponding transcriptional changes are visualised in Figure 6.2. Gene expression is displayed in heatmaps with upregulation (i.e., higher in FCS shown in red) and downregulation (i.e., higher in RPMI shown in blue) for each of our two strains at time points 4h and 24h.

From the transcriptional activity it is possible to see corresponding changes that align with our metabolomics data (Figure 5.2). Notably biosynthesis of unsaturated fatty acids displayed increased gene expression in our LBF at 24h in FCS. This same pattern is not observed in the HBF strain. Many of these distinguishing genes in LBF were also observed in the propanoate metabolic pathway. Fatty acid metabolism has been previously shown to modulate morphogenesis in *C. albicans* (Noverr and Huffnagle, 2004). Our observations here would indicate that the observed phenotypic switching or increased morphogenesis that we observe in our LBF is through the utilisation of these pathways. From our overlayed transcriptional expression in Figure 5.2, we can see LBF are alternating to beta-oxidation of fatty acids. The genes within the Glycine, serine and threonine pathway SER33, GCV2, GCV1 are upregulated in in FCS to a greater degree than HBF at 4 hours additionally IFG3 and SHM2 are also upregulated in FCS in HBF but to a lesser extent. Pyruvate metabolism genes show a similar pattern at 4h in both strains apart from LYS22 which encodes homocitrate synthase activity. ALD5 is upregulated in LBF FCS at 24h and not HBF as is the putative gene C7_02010C. ALD5 is aldehyde dehydrogenase and C7_02010C shares homology with aldehyde dehydrogenase according to the CGD, which are involved in the biosynthesis of acetate. Both are also identified in the pantothenate and CoA biosynthesis pathways, as well as the lysine degradation pathways. Glycerolipid and glycerophospholipid showed differential activity in

the metabolomics data glycerolipid metabolites showed high (secreted activity) in FCS and high (spent activity) in the RPMI. Visualised higher levels of gene expression are observed in some FCS with very high levels in certain genes in the HBF. Through overlaying our transcripts on our metabolic pathways, it is possible to discern that there are corresponding features within the transcriptomics as shown in Figure 5.2. And some patterns such as the gene expression in biosynthesis of unsaturated fatty acids are easier to discern than others.

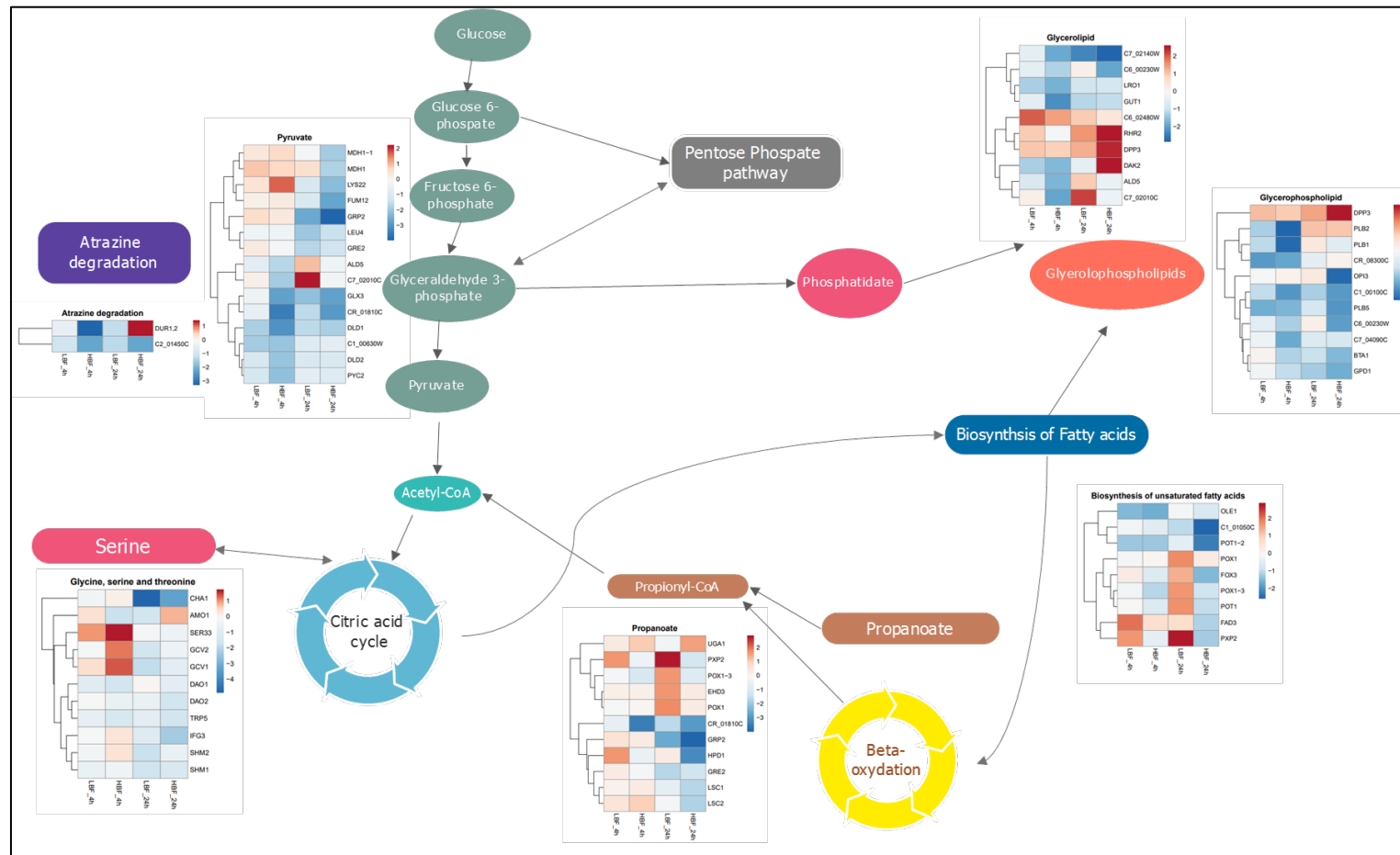


Figure 5.2 **Differential expression of genes associated with key metabolic pathways.** Differential expression of *C. albicans* transcripts in FCS vs Serum was performed between LBF and HBF at 4h and 24h. Those differentially expressed genes that were differentially expressed and that also belonged to a pathway that was determined to be significant by PAPI analysis based upon differences in activity score between our sample groups according to FDR corrected ANOVA. Pathways filtered first on significance within the PAPI analysis then pathways with significantly DE genes from at least one condition are displayed. Leaving genes involved in pathways that. Log2 fold change RPMI compared to Serum for each strain at both 4 and 24h are represented within the heatmaps within each pathway shown.

5.4.2 Pathway models

Pathway driven models are based on *a priori* knowledge on biological databases that are maintained either generically or for individual species. These databases include, but are not limited to, the KEGG, Reactome, wikiPathways, MSigDb and Gene Ontology databases (Kanehisa et al., 2015, Jassal et al., 2020, Slenter et al., 2018). Methods of functional scoring or enrichment have become prominent, especially with classification of genes within transcriptome studies, in which pathways are scored or enriched dependent upon how represented they are. Gene set enrichment analysis and over representation-based analysis have become popular methods to identify the most represented functional classes from these databases. The hypergeometric distribution of identifying overrepresented classes utilises a list of genes of interest typically those that are differentially expressed. The hypergeometric distribution is utilised to assess whether that set of genes are significantly associated with a specific pathway or classifier. In contrast, the GSEA methodology utilises the entire list of features without any prior filtering or selection. GSEA uses the aggregates of per gene or features within the set-in order to address the limitation of ORA in which identification of small but coordinated changes are missed (Subramanian et al., 2005). By this method pathways that have small but consistent changes across the entire gene set are detected. This method also requires the gene rank in contrast to ORA. The gene rank based upon the expression score commonly the log fold change of a gene within gene list. These techniques are no longer limited to gene expression and ORA and metabolite set enrichment analysis (MSEA) are utilised within the field of metabolomics. ORA for metabolomics is available through numerous bioinformatic tools including IMPaLA, MBRole, MetaboAnalyst, MetExplore and PiMP. MSEA is also available through a number of these tools including the widely used MetaboAnalyst (Gloaguen et al., 2017, Cottret et al., 2018, Chong et al., 2018a, Kamburov et al., 2011).

The pathway approach to many single omic studies has been a preferable way to approach omics data with a large number of features that can be difficult to interrogate and interpret at an individual level. Pathway and functional classification of gene, protein, metabolite and other molecular level features considered in regard to their overall function within a cell or organism have been chosen by many researchers to interpret the phenotypic changes that they observe (Cruickshank-Quinn et al., 2018). Pathway analysis methods have been

expanded to the integration of more than metabolomics. Many of the software tools noted including IMPaLA, MetaboAnalyst, Metexplore have either been created or expanded to incorporate the integration of transcriptomic data (Chong et al., 2018a, Cottret et al., 2018, Kamburov et al., 2011). The specifically address IMPaLA for example was first utilised in 2011 to integrate transcriptome and metabolism in consensus pathway-based approach. The approach is essentially an extension of the ORA using Fishers method to consider pathway overrepresentation based upon the joint probabilities from both the transcriptome and metabolome feature lists (Kamburov et al., 2011). It was noted that due to the necessity of robustly annotated databases of pathways for both genes and metabolomics there exists a limitation that this technique cannot identify any novel functions. It could also be added that this study was a human biological study in which the technique was tested and there are other considerations when applying integration to other organisms. Specific microbes suffer in comparison to human and mouse studies in that they are less well studied. This means that limitation that exist in pathway-based approaches are amplified for other organisms and microorganisms.

This approach has gained popularity and is included within other popular and visualisation packages. These include PaintOmics 3 which has a specific focus on integration of data from multiple omics sources including proteomics and utilises the KEGG database for feature classification. MetaboAnalyst which is widely popular and has become a “one stop shop” for metabolomics analysis incorporates many of the popular tools for single omics and now the ability to incorporate transcriptomic data. Joint pathway enrichment/over representation methods are utilised by both tools in this specific application, and both offer a number of methods for visualising the data. Limitations exist for individual species across many of the integration and single omics tools in OR. It is possible to map identifiers to the model yeast *S. cerevisiae* for analysis of *C. albicans* and other yeasts. Availability of different species databases within these tools is increasing and this will inevitably continue to grow as omics and analysis tools become more utilised within microbial research.

An important consideration is that when utilising RNA-Sequencing you measure the total mRNA and therefore all features are represented. Metabolomics data depending on the instrument and method utilised can influence the metabolites that are represented

(Bhinderwala et al., 2018). Metabolites are more likely or easier to find dependant on whether NMR, GC-MS or LC-MS is used in acquiring metabolomic data.

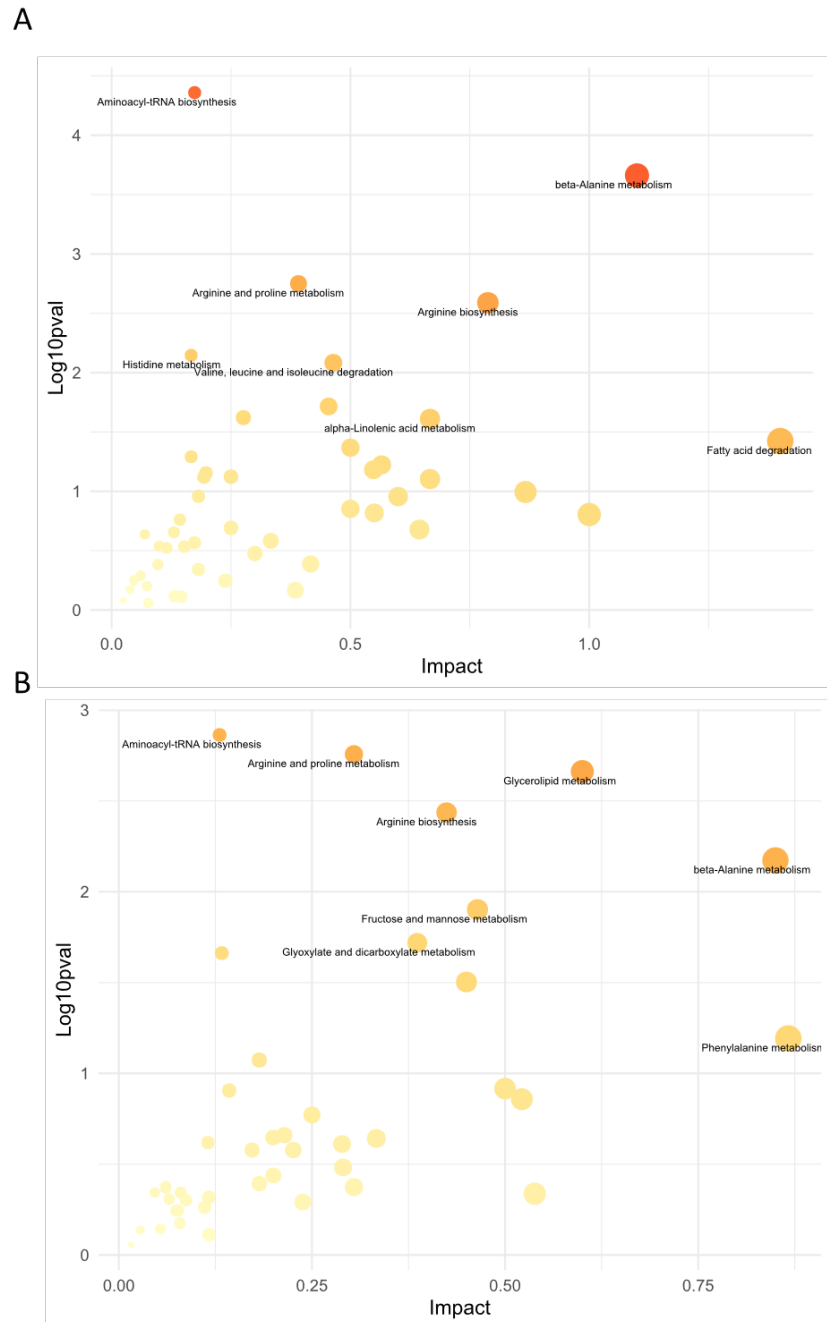


Figure 5.3 Integrated pathway enrichment from *C. albicans* transcripts and metabolomes. Upregulated genes in FCS compared to RPMI media after growth of biofilms in for 24h were mapping of IDs to *S. cerevisiae* orthologs, before being submitted for joint pathway analysis through MetaboAnalyst. Metabolites that were also in differential abundance between *C. albicans* isolates grown in FCS media compared RPMI at 24h were also submitted for joint pathway over representation analysis. From this the joint inverse p-value ($-\text{Log}_{10}\text{pval}$) and pathway impact from central and highly important pathways. This was performed for both LBF (A) and HBF (B) isolates comparing pathways impacted in FCS grown cells in a joint metabolome-transcriptome analysis. The pathway impact is plotted on the x with the inverse joint log10 p value plotted on the y. The node size increases with pathway impact.

The method utilised here was the method of pathway enrichment utilised by the widely used MetaboAnalyst software package. This software supports the overlapping of pathways from multiple organisms which currently does not support *C. albicans*, but it does support *S. cerevisiae*. *S. cerevisiae* is a commonly used model organism from which a lot of the homology information of *C. albicans* is drawn (Skrzypek et al., 2016). *Candida* gene IDs from the CGD were first mapped to their corresponding *Saccharomyces* corresponding systematic name. Upregulated genes from HBF at 24 in FCS and LBF at 24h in FCS relative to RPMI were selected based upon our previous cut-offs. Like ORA performed on individual datasets this list of “important” features is what are submitted. A corresponding list of metabolites assigned to their KEGG identifiers, from our metabolomics analysis were also submitted for pathway enrichment analysis. In total two enrichment analysis were performed one for LBF and one for HBF both at 24h (Figure 5.3). Metabolites and genes are then integrated either by combining queries. Combining queries pools both metabolites and transcripts. Or by performing two separate over representation analysis and combining p-values. Pathway-level p-value joining was chosen for our data. Pathway level weights the gene and metabolite data based on how many genes or metabolites are in the corresponding pathways. Pathway level also allows for pathways to be enriched even if there are only features available from one data set (either genes or metabolites). We observed that amino-acyl-tRNA biosynthesis was the most significantly upregulated according to false detection rate (FDR) adjusted p-value in both LBF and HBF. It is important to note that aminoacyl tRNA synthesis has only one enzyme within this pathway and therefore it is significantly overrepresented within our data but has a lower bar for enrichment due to the few features. However, this is indicated by the low pathway impact score received by this pathway despite its significance value. KEGG and other databases include aminoacyl tRNA synthesis within their pathways however it is perhaps should not be considered as such. Pathway analysis attempts to account for pathways with low smaller or larger numbers of features such as in the case of hypergeometric analysis.

This significance was driven by mapped metabolites and not transcripts. Arginine and proline were over-enriched in both LBF and HBF as was arginine biosynthesis. In LBF beta-alanine was also enriched significantly but not in HBF. In HBF the Glycerolipid metabolism was enriched but not in LBF. Joint pathway scores and metabolite/gene numbers are available in Appendix IV.

5.4.3 Statistical and multivariate methods

Multivariate based methods or dimensionality reduction-based approaches are widely used within bioinformatics. They are often applied to single omics data, such as the commonly used principal component analysis (PCA), to reduce the data and identify the largest variation which exists within the feature set (Marini and Binder, 2019). PCA seeks the directions that account for the largest variation and therefore can visualise natural separation in the data. Separation that is due to the overall variation in the data is determined by PCA and because of this is classified as an unsupervised dimensionality reduction technique (Stein-O'Brien et al., 2018).

Several multivariate methods in recent years have been proposed for integrating omics datasets with a number of variations of these techniques appearing within the literature. Notably partial least squares (PLS) regression which aims to maximise the covariance between the datasets (e.g. Transcriptome and Metabolome (Wold, 1973). Also, Canonical Correlation Analysis (CCA) which aims to maximise correlation rather than covariance. Both have been utilised as methods of dimensionality reduction, classification and in variable discrimination (Wold, 1973, González et al., 2012).

PLS and CCA in contrast to PCA can correlate the information from one matrix [x] to the information in another [y]. In omics x and y are our feature tables from two omics data experiments. PLS is perhaps the most widely used multivariate technique for data integration of omics data in particular integration of metabolomic-transcriptomic data (Worley and Powers, 2013). Due to its capabilities in handling data with high levels of collinearity and when trying to identify aspects of one dataset that predict features of another or covariance such as in the case of metabolomic and transcriptomic data (González et al., 2012). PLS is long standing multivariate tool which pre-dates omics technologies however was applied to omics data sets as early as 2004 (Griffin et al., 2004). Sparse PLS in transcriptomic-metabolomic integration is novel variation of the PLS method which arose specifically to address the selection of variables for biological interpretation through lasso penalization (KA et al., 2008). This method allows for the variable selection and classification allowing for the use of PLS which allows it to resolve the classification problems. PLS discriminant analysis (PLS-DA) has also been shown to be another effective variation of PLS capable of performing

classification and discrimination on similar omics sets (Boccard and Rutledge, 2013). The sparse PLS-DA described by Lê Cao et al is able to identify discriminatory variables given a matrix of features and their classes (Cao et al., 2011).

Similar to pathway analysis, a field of data integration by multivariate methods has arisen and now many of these methods are becoming available in software packages. Many of these are biostatistical packages written in an implanted through the R programming language. These include *integrOmics*, *MixOmics*, *OmicsPLS* and *O2PLS* in the *Siumca* software package (Rohart et al., 2017, KA et al., 2009, Bouhaddani et al., 2018). *MixOmics* offers several multivariate methods both supervised and unsupervised for single, dual and multiple integration models (Rohart et al., 2017).

We utilised *MixOmics* to assess multivariate integration upon multiple omics datasets. We utilised a supervised discriminatory analysis to identify features which discriminate between our classes. Our classes provided were the two time points and the two medias. We did not classify the samples in terms of their strain due to low replicate number and to primarily identify covariate features in the Serum vs RPMI. The *MixOmics* contains several analysis methods some which are common and others which are novel within the *MixOmics* package. Supervised classification of multiple omics data is provided by their *block.plsda* and *sPLS-DA* methods. The aim of which is to identify co-expressed variables within our dataset which explain our classes as described. A caveat to this approach with our dataset these integration methods are described as highly sensitive to differences in omics platforms even in the case of measurements on the same samples. Due to our repeat measures study design in which we prepared to different batches of samples, one for metabolomics and one for transcriptomics. This batch effect is something that we are unable to control for across different omics technologies and therefore we utilised multivariate integration as “fishing” or hypothesis driving tool and therefore do not intend to draw firm conclusions from them.

For integrating samples, we use *block.splsda* function from within *MixOmics* we utilised log relative normalised transcripts from DESeq2 and the log media normalised metabolites. Metabolites were filtered to leave only identified metabolites within the data set and reduced the metabolomics data set so that only the matching LBF and HBF strain used in both assays

remained. After our matched datasets were created the *block.splsda* uses a multiblock method for data integration. Where the blocks are the data types organised by matching sample identifiers. sPLS-DA combines PLS-DA with Lasso penalisation for selecting the most discriminant variables. Samples were assigned as time and media and strain type was ignored. Our hypothesis was to maximise the discriminant variables between FCS and RPMI. sPLS-DA multiblock has the dual objective of maximising the co-variance within the blocks of omics (i.e., metabolomic and transcriptomic) whilst discriminating between our groups. The correct number of components was tuned based on the balanced error rate which was deemed to be 2 components. Tuning parameters were used to determine the number of metabolites and transcripts needed for discrimination on each component this was 20,20 for transcripts and 10,10 for metabolites. From our block plots shown in Figure 5.4A our first component discriminated between time and our second component by media with our most covariate variables Figure 5.4B. Discriminant features and their contributions to component 2 are shown in Figure 5.5A and the correlation between these variables are shown in the circus plot 5.5B. We identified a subset of our identified metabolites that discriminant between *C. albicans* grown in FCS vs RPMI and secondly covariate genes that discriminate between these two biofilms. VPS28 is an important discriminatory feature in FCS involved in the proteolytic activation of Rim101, a pH alkaline response promoter if hyphal formation (Figure 5.4). DAD2 also an identified discriminator is involved in microtubule binding subcomplex DASH required for proper chromosome segregation (Figure 5.4). STRING analysis which identifies protein-protein interactions based on function, homology, publication history and many other identifiable relationships did not identify any common links between our subset of transcripts (Szkłarczyk et al., 2019).

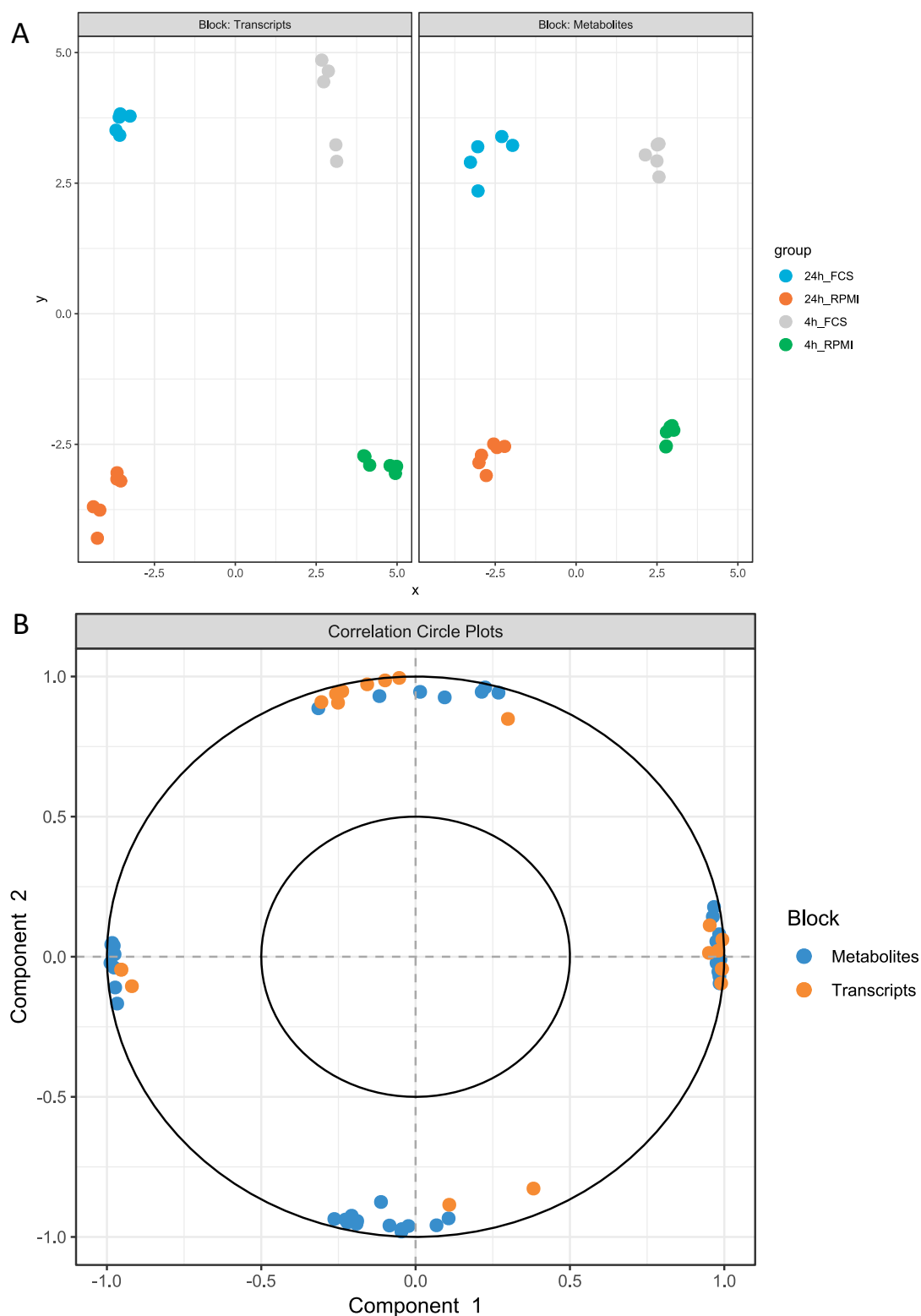


Figure 5.4 **Sparse partial least squares discriminant analysis (sPLS-DA) of *C. albicans* isolates in the presence and absence of serum.** The log2 relative to the media control normalised metabolome matrix and the DESeq2 normalised transcript matrix were submitted to sPLS-DA in the mixOmics package. Data projection of the transcriptome data and the metabolome data based upon the most discriminant features are shown (A). Sample groups of timepoint and media that isolates are grown in are colourised. The correlation between the most discriminant metabolites and transcripts are shown (B). Metabolic and transcriptomic features are colourise.

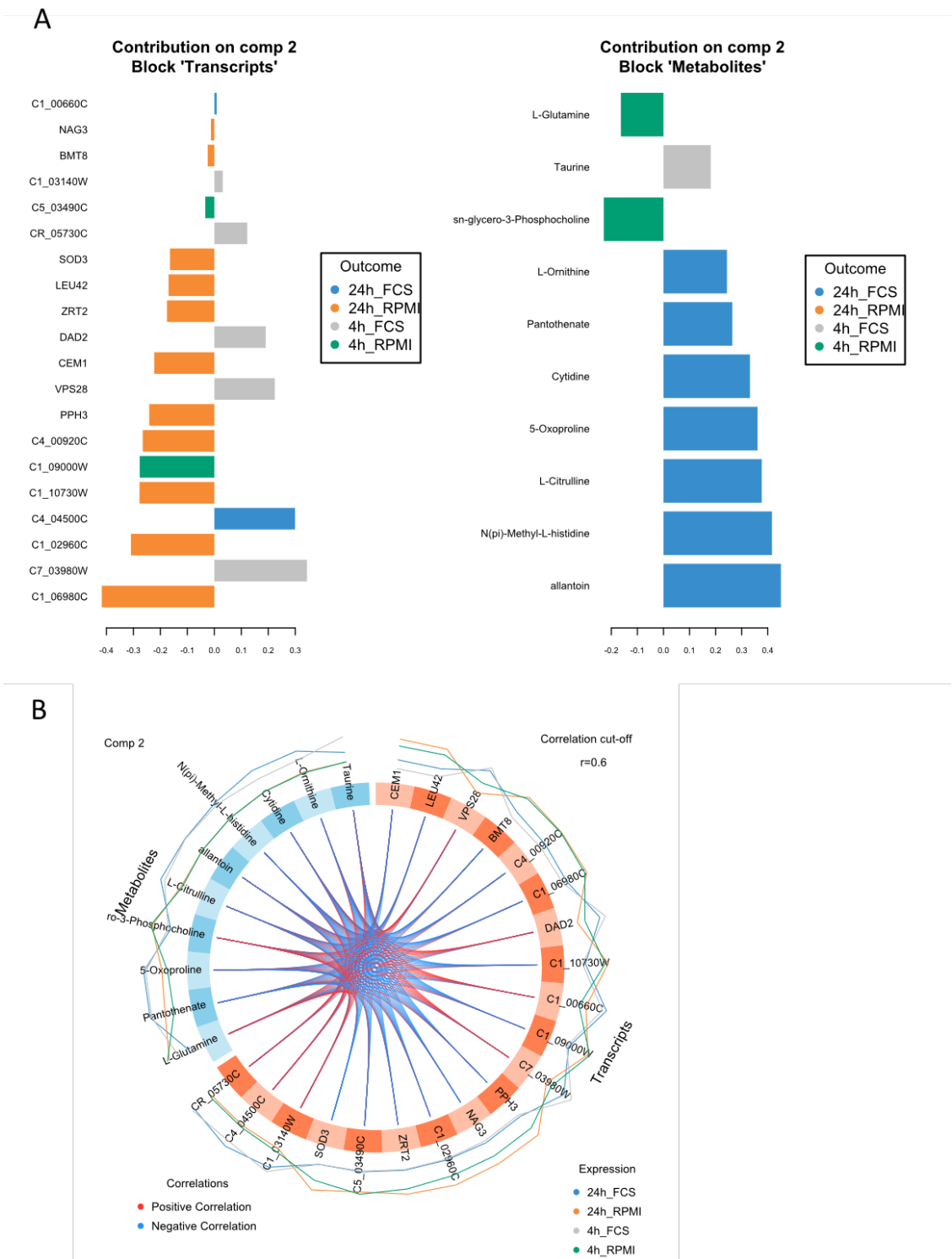


Figure 5.5 Important features and correlations between metabolites and transcripts. Loadings plot of discriminant features on component 2 (A). The loading weight for transcripts and metabolites which discriminate most highly on component two are displayed in a bar plot. Colour of the bar is dependent on the sample in which the feature (transcript or metabolite) is most abundant. Correlations of discriminant metabolites and transcripts (B). Circos plot displays the correlation between all our discriminatory variables on component 2. Component 2 shows the largest separation between samples grown in FCS compared to RPMI. Lines between metabolites and transcripts indicate correlation and coloured either red for +ve or blue for +ve correlation.

5.4.4 Data Integration Conclusions

Through metabolomic and transcriptomic profiling we have identified several key pathways that are differentially changing between *C. albicans* in serum vs RPMI. Additionally, we have been able to identify metabolic networks which are being reprogrammed in otherwise biofilm deficient strains to induce them to form biofilms and undergo morphogenesis. Fatty acid biosynthesis seems to be key appearing in transcriptomic and metabolomic data alike as an important feature in the metabolomic adaptation in serum.

By integrating the two data sets we have identified important features which were not apparent when considering each individual data set alone. To date integrated metabolomics and transcriptomics have rarely been performed on *in vitro* biofilm models. There are limitations that exist within the data which have been highlighted throughout. Using both knowledge based and multivariate methods however it is possible to apply these analytical techniques to *in vitro* to determine both underlying pathways and important “signatures” within the models. As our bioinformatic databases on non-model organisms continues to grow as with the growth of omics technologies and it is these continued advancements will continue to help us decipher biofilms *in vitro*.

5.5 Final discussion and Conclusions

C. albicans biofilms are increasingly being recognised as important contributors to clinical infections, where historically they had been dismissed. These biofilms compared to planktonic cells are more resilient and drug resistant. Understanding the mechanisms of biofilm formation have become a key focus in *Candida* research. Through reviewing the literature and careful experimentation, with two distinct phenotypes, in biofilm inducing conditions we have begun to decipher the pathways driving this phenomenon in clinical isolates.

Within chapters 3, 4 and above, we have shown that within different clinical isolates of *C. albicans*, their phenotypic behaviour is distinct, which is governed by their transcriptomic and metabolomic profiles. Further to this even under the same stimulus, in our case serum, their metabolome and transcriptomic response is not completely parallel. These are observations that would not be made with traditional *in vitro* that focuses on type strains. Serum is just

one stimulus and *C. albicans* biofilm formation is differentially affected by number of known stimulus including pH, host factors, nutrients, and bacterial interactions. Biofilm formation is complex, and *C. albicans* is known to have a regulatory network of 6 genes that in turn regulate 1000 genes responsible, or linked to, hyphal formation (Nobile et al., 2012). We and others have also observed this to be heterogenous, especially within clinical strains. Another layer of complexity is the interactions of differentially “biofilm-ready” strains and their interactions with these promoters such as serum. As we have highlighted, the response of different *C. albicans* isolates to serum is not consistent and regulation of different metabolic networks to achieve hyphal and biofilm formation is strain dependant (Rajendran et al., 2016a, Rajendran et al., 2016d). LBF do not respond to RPMI 1640, which typically triggers hyphal and biofilm formation due to the copious amounts of glucose (Kucharíková et al., 2011), a response that is observed in well characterised type strains and many of our clinical isolates (Weerasekera et al., 2016, Rajendran et al., 2016d). Biofilm formation through glucose sensing in *C. albicans* is modulated through three key networks glucose repression, the sugar receptor repressor (SRR), and adenylate cyclase networks. Two of these cascades, the SRR and adenylate cyclase are implicated in cellular morphogenesis. LBF strains do not respond to this stimulus in the same way and remain unadhered and predominantly as yeast. However, as we have demonstrated they are not biofilm defective, and additional stimuli are able to induce alternative metabolic reprogramming to induce this phenotype. In serum the LBF can utilise the fatty acid metabolic pathway to induce this phenotype. LBF and HBF seemingly differed more in their transcriptional activity than in their metabolic. A limitation of our study design perhaps that we did not have the intracellular metabolic activity to interrogate alongside our extracellular data. As indicated previously within the thesis, improved metabolomics techniques such as ms/ms and metabolic flux would also perhaps demonstrate with greater resolution the activation of metabolic networks of our different strains under the different stressors.

Throughout this work LBF and HBF isolates were compared against each other and were grown in either FCS or RPMI. RPMI was considered the baseline or control media with the supplementation of FCS to the media being our test condition. Similarly, HBF was our point of comparison with which we compared the LBF isolates. This was due to the larger and much more stark effect of FCS on LBF as shown in chapter 3. In which the non-biofilm formers

morphogenically switch to biofilm formers in FCS supplemented media. FCS is poorly defined and poorly controlled media, particularly when compared to synthetically made medias such as RPMI. This introduces lots of variability which cannot be controlled. Further experimentation would benefit from introduction of metabolites, metals and vitamins found in FCS in known quantities to provide more reproducible and better controlled experimentation. Such as the introduction of individual amino acids or metal ions such as zinc to minimal media. In this fashion we could identify the individual effects of metabolites and potential biofilm influencers without confounding factors and variability from poorly defined medias. These are avenues of research that could be taken forward from the work presented here.

Within this body of work and throughout much of the work within Oral Sciences group we have opted to use clinical isolates in the place of laboratory type strains. Type strains, despite being well defined, are unrepresentative and do not model the heterogeneity that is seen in clinical samples. This heterogeneity amongst clinical isolates is well highlighted in the Scottish candidaemia study from which these isolates were received (Rajendran et al., 2016d). However, comparison to a type of strain control may have been beneficial when identifying shared and differing responses, due to the literature on biofilm formation strains such as SC5314 being much larger. Another limitation is that throughout many of our comparisons we only used a small number of isolates, such as those within the RNA-Seq work in chapter 3 where we used 1 HBF and 1 LBF. We used these to represent our two phenotypes to compare the effects of FCS on Low vs High Biofilm formers. However, in the light of how much heterogeneity and biofilm forming variation there is between strains, as highlighted by the Candidemia study, other methodologies may have been more appropriate. For example, rather than a 1 vs 1 controlled experiment with technical replication, profiling of a larger cohort of isolates may have yielded further information. By comparing a larger number of isolates, we would have been able to observe trends and correlations in features from metabolomic, transcriptomic and proteomic with the observed phenotype. This study design would have been less biased towards select isolates that show extremes of the two isolates and discerned more suitable differences that are related to phenotype.

We have previously described metabolomics and transcriptomic experiments as hypothesis driving methodologies in contrast to hypothesis testing experiments. A limitation of the work presented in this thesis is the lack of confirmatory experimentation. For example, work using minimal media with metabolites incrementally added would have confirmed some of our findings. Addition of zinc to the media to confirm the downregulation of these genes for example. Further work to elucidate the presence of arachidonic acid such as confirmatory assays available in the form of enzyme-linked immunosorbent assays (ELISA). The lack of these assays are due to a time and lab access restriction but would be interesting avenues of research for the Oral Sciences research group. Similarly, we saw amino acid metabolism and fatty acid metabolism changes. Experimentation with different phenotypic isolates of *Candida* in the presence and absence of amino acids such as proline, arginine and serine could yield interesting results. Simple biofilm assays such as crystal violet could inform whether amino acid presence is able to induce biofilm formation in low or intermediate biofilm formers. Similarly, it is common to confirm RNA-Seq findings with qPCR of select genes of interest. The aim being to check that qPCR findings match with our high throughput observations. This is commonly performed using RNA that is retrieved from samples that have had all the same experimental steps as the RNA-Seq data. There has been arguments made that the information that this adds is low and that RNA-Seq is robust enough that this is no longer necessary (Coenye, 2021). Validating by qPCR has potentially been inherited from microarrays which had higher error and bias compared to RNA-Seq and RNA-Seq by comparison shows low concordant results with only around 1.8% of genes (Everaert et al., 2017). Additionally, the use of small number of genes to validate a much larger dataset may seem unintuitive. As is discussed in the review from 2021 the lack of correlation with a single gene through validation does not speak to the validity of the remainder of the dataset (Coenye, 2021). However, due to our collection of blood stream isolates being so large with >100 categorised isolates then qPCR could possibly add to our findings. Taking a larger and more varied selection of *C. albicans* isolates we could have profiled the expression of genes of interest, such as the SAPs which were upregulated in FCS, zinc transporters which were up in RPMI or the ATO gene family which were specific to the LBF in FCS, in a much larger selection of isolates. This would have allowed us to see if the differential expression of these genes was consistent across phenotypically similar strains. This adds to the previous point that a simpler

RNA-Seq study design on a much larger selection of biologically varying *C. albicans* isolates would similarly add more power to the study design.

Throughout this thesis we have discussed multi-omics and focused on transcriptomic and metabolomic analysis. There are many other omics, including lipidomic and genomics. There are also many variations of some of the techniques used as we highlighted in the metabolomics section. Notably missing from this work is the inclusion of genomic data. A genomic profile of all the isolates from the Scottish candidaemia study would be the logical progression. Whole genome sequencing by Illumina or similar platforms followed by assembly of genomes and identification of variants, either copy number or single nucleotide variants (SNV). This is future work that will be of great importance, through using NGS to genotypically understand any phenotypic and clinical variation within these strains. This work would answer questions both about variation that determines our observed phenotypes *in vitro* and variation related to our epidemiological findings (Rajendran et al., 2016d). These isolates were highly curated with both clinical data and antifungal resistance and sensitivity data. Links to the genomic data could also provide insights into the heterogeneity in antifungal susceptibility and resistance that was observed between these isolates. Functional genomics is an effective tool at linking the genome to phenotype and pathogenicity/virulence related functions in yeasts and fungi (Farrer and Fisher, 2017).

Within the literature review we also highlighted the importance of transcriptional control on the much larger network of biofilm and hyphal genes. These were important findings that were first discovered by Nobile and colleagues in 2012 (Nobile et al., 2012). Transcription factor activity through these cascades could therefore have been profiled using phosphoproteomics. Phosphoproteomics identifies active pathways through mass spectrometry phosphoprotein analysis (Dudley and Bond, 2014). This methodology has previously been shown to be effective in tracking the phosphorylation of proteins involved in pathogenesis in *C. albicans* and was applied to identify the activity of protein kinases during hyphal elongation (Ghorai et al., 2018). It would be interesting therefore to see if there were any changes in the dynamics of these hyphal regulating kinase cascades between high and low forming isolates. To our knowledge this would be the first study looking at the differences

in protein kinase and kinase interaction with phosphoproteins between clinically relevant isolates.

Considering omics approaches as hypothesis drivers and not hypothesis testers, however, this body of work opens several avenues of investigation. Further work would hope to expand on the inducibility of biofilm formation from other known inducers. It is also assumed that HBF are more tolerant to antimicrobials, and to date little is known about whether this resistance is shared by LBF that have been induced. Because hyphal formation and biofilm formation is inducible it is perhaps also possible to hypothesize that these conditions exist within the host. Novel antifungal regimens that consider alternative metabolic pathways may also yield improved antifungal sensitivity *in vivo*. As discussed, we have only discussed one stimulus out of the many that are known to induce heterogeneity. Anecdotally, work not presented in this thesis show that certain bacteria are able to induce this biofilm phenotype. This indicates that biofilm heterogeneity is more complex, and perhaps these LBF strains are more adapted to co-exist mutually or synergistically with bacterial species. It has been shown that the relationship with bacteria is a growing trend and within the oral cavity we additionally demonstrated that fungi do impact the microbial community. *Candida's* role within multi-species biofilms is still needed to be understood fully and how both *Candida* and bacteria interact with each other. This is particularly true in the oral cavity as we have identified from the literature over the last 20 years and our own microbiome studies. The studies outlined here lead the way for further omics profiling of increasingly complex interkingdom communities. It is our belief that these studies need to be performed and the thought of what omic studies could tell us about interkingdom biofilm communities and their interactions is tantalising.

References

2006. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res*, 34, D322-6.
2020. RefSeq [Online]. Available: <https://www.ncbi.nlm.nih.gov/refseq/> [Accessed November 2020].
- AAS, J. A., PASTER, B. J., STOKES, L. N., OLSEN, I. & DEWHIRST, F. E. 2005. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol*, 43, 5721-32.
- ABUSLEME, L., DUPUY, A. K., DUTZAN, N., SILVA, N., BURLESON, J. A., STRAUSBAUGH, L. D., GAMONAL, J. & DIAZ, P. I. 2013. The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME J*, 7, 1016-25.
- AGGIO, R. B., RUGGIERO, K. & VILLAS-BÔAS, S. G. 2010. Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity. *Bioinformatics*, 26, 2969-76.
- AL-HEBSHI, N. N., BARANIYA, D., CHEN, T., HILL, J., PURI, S., TELLEZ, M., HASAN, N. A., COLWELL, R. R. & ISMAIL, A. 2019. Metagenome sequencing-based strain-level and functional characterization of supragingival microbiome associated with dental caries in children. *Journal of Oral Microbiology*, 11, 1557986.
- AL MUBARAK, S., ROBERT, A. A., BASKARADOSS, J. K., AL-ZOMAN, K., AL SOHAIL, A., ALSUWYED, A. & CIANCIO, S. 2013. The prevalence of oral Candida infections in periodontitis patients with type 2 diabetes mellitus. *J Infect Public Health*, 6, 296-301.
- ALALWAN, H., NILE, C. J., RAJENDRAN, R., MCKERLIE, R., REYNOLDS, P., GADEGAARD, N. & RAMAGE, G. 2018. Nanoimprinting of biomedical polymers reduces candidal physical adhesion. *Nanomedicine*, 14, 1045-1049.
- ALIZADEH, M., KOLECKA, A., BOEKHOUT, T., ZARRINFAR, H., GHANBARI NAHZAG, M. A., BADIEE, P., REZAEI-MATEHKOLAEI, A., FATA, A., DOLATABADI, S. & NAJAFZADEH, M. J. 2017. Identification of Candida species isolated from vulvovaginitis using matrix assisted laser desorption ionization-time of flight mass spectrometry. *Curr Med Mycol*, 3, 21-25.
- ALLEN, F. & LOCKER, D. 2002. A modified short version of the oral health impact profile for assessing health-related quality of life in edentulous adults. *Int J Prosthodont*, 15, 446-50.
- ALNUAIMI, A. D., O'BRIEN-SIMPSON, N. M., REYNOLDS, E. C. & MCCULLOUGH, M. J. 2013. Clinical isolates and laboratory reference Candida species and strains have varying abilities to form biofilms. *FEMS Yeast Res*, 13, 689-99.
- ALVES, C. T., WEI, X.-Q., SILVA, S., AZEREDO, J., HENRIQUES, M. & WILLIAMS, D. W. 2014. *Candida albicans* promotes invasion and colonisation of *Candida glabrata* in a reconstituted human vaginal epithelium. *Journal of Infection*, 69, 396-407.
- AMIR, A., MCDONALD, D., NAVAS-MOLINA, J. A., KOPYLOVA, E., MORTON, J. T., ZECH XU, Z., KIGHTLEY, E. P., THOMPSON, L. R., HYDE, E. R., GONZALEZ, A. & KNIGHT, R. 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*, 2, e00191-16.
- ANDERS, S., PYL, P. T. & HUBER, W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 31, 166-169.

References

- ANDES, D. R., SAFDAR, N., BADDLEY, J. W., PLAYFORD, G., REBOLI, A. C., REX, J. H., SOBEL, J. D., PAPPAS, P. G., KULLBERG, B. J. & MYCOSES STUDY, G. 2012. Impact of treatment strategy on outcomes in patients with candidemia and other forms of invasive candidiasis: a patient-level quantitative review of randomized trials. *Clin Infect Dis*, 54, 1110-22.
- AOKI, W., UEDA, T., TATSUKAMI, Y., KITAHARA, N., MORISAKA, H., KURODA, K. & UEDA, M. 2013. Time-course proteomic profile of *Candida albicans* during adaptation to a fetal serum. *Pathogens and Disease*, 67, 67-75.
- APWEILER, R., BAIROCH, A., WU, C. H., BARKER, W. C., BOECKMANN, B., FERRO, S., GASTEIGER, E., HUANG, H., LOPEZ, R., MAGRANE, M., MARTIN, M. J., NATALE, D. A., O'DONOVAN, C., REDASCHI, N. & YEH, L. S. 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*, 32, D115-9.
- ARITA, G. S., MENEGUELLO, J. E., SAKITA, K. M., FARIA, D. R., PILAU, E. J., GHIRALDI-LOPES, L. D., CAMPANERUT-SÁ, P. A. Z., KIOSHIMA, É. S., BONFIM-MENDONÇA, P. D. S. & SVIDZINSKI, T. I. E. 2019. Serial Systemic *Candida albicans* Infection Highlighted by Proteomics. *Frontiers in Cellular and Infection Microbiology*, 9.
- AVILA, M., OJCIUS, D. M. & YILMAZ, O. 2009. The oral microbiota: living with a permanent guest. *DNA Cell Biol*, 28, 405-11.
- AZEREDO, J., AZEVEDO, N. F., BRIANDET, R., CERCA, N., COENYE, T., COSTA, A. R., DESVAUX, M., DI BONAVENTURA, G., HEBRAUD, M., JAGLIC, Z., KACANIOVA, M., KNOCHER, S., LOURENCO, A., MERGULHAO, F., MEYER, R. L., NYCHAS, G., SIMOES, M., TRESSE, O. & STERNBERG, C. 2017. Critical review on biofilm methods. *Crit Rev Microbiol*, 43, 313-351.
- BACHTIAR, E. W., BACHTIAR, B. M., JAROSZ, L. M., AMIR, L. R., SUNARTO, H., GANIN, H., MEIJLER, M. M. & KROM, B. P. 2014. AI-2 of *Aggregatibacter actinomycetemcomitans* inhibits *Candida albicans* biofilm formation. *Front Cell Infect Microbiol*, 4, 94.
- BAKER, J. L., MORTON, J. T., DINIS, M., ALVEREZ, R., TRAN, N. C., KNIGHT, R. & EDLUND, A. 2019. Deep metagenomics examines the oral microbiome during dental caries, revealing novel taxa and co-occurrences with host molecules. *bioRxiv*, 804443.
- BALVOČIŪTĖ, M. & HUSON, D. H. 2017. SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? *BMC Genomics*, 18, 114.
- BAMFORD, C. V., D'MELLO, A., NOBBS, A. H., DUTTON, L. C., VICKERMAN, M. M. & JENKINSON, H. F. 2009. *Streptococcus gordonii* modulates *Candida albicans* biofilm formation through intergeneric communication. *Infect Immun*, 77, 3696-704.
- BASTIAAN, R. J. 1976. Denture sore mouth. Aetiological aspects and treatment. *Aust Dent J*, 21, 375-82.
- BHARTI, R. & GRIMM, D. G. 2019. Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics*.
- BHINDERWALA, F., WASE, N., DIRUSSO, C. & POWERS, R. 2018. Combining Mass Spectrometry and NMR Improves Metabolite Detection and Annotation. *Journal of proteome research*, 17, 4017-4022.
- BIEN, J., SOKOLOVA, O. & BOZKO, P. 2011. Characterization of Virulence Factors of *Staphylococcus aureus*: Novel Function of Known Virulence Factors That Are Implicated in Activation of Airway Epithelial Proinflammatory Response. *J Pathog*, 2011, 601905.

References

- BILHAN, H., SULUN, T., ERKOSE, G., KURT, H., ERTURAN, Z., KUTAY, O. & BILGIN, T. 2008. The role of *Candida albicans* hyphae and *Lactobacillus* in denture-related stomatitis. *Clinical Oral Investigations*, 13, 363.
- BINDEA, G., MLECNİK, B., HACKL, H., CHAROENTONG, P., TOSOLINI, M., KIRILOVSKY, A., FRIDMAN, W.-H., PAGÈS, F., TRAJANOSKI, Z. & GALON, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)*, 25, 1091-1093.
- BINK, A., VANDENBOSCH, D., COENYE, T., NELIS, H., CAMMUE, B. P. A. & THEVISSSEN, K. 2011. Superoxide dismutases are involved in *Candida albicans* biofilm persistence against miconazole. *Antimicrobial agents and chemotherapy*, 55, 4033-4037.
- BITAR, I., KHALAF, R. A., HARASTANI, H. & TOKAJIAN, S. 2014. Identification, typing, antifungal resistance profile, and biofilm formation of *Candida albicans* isolates from Lebanese hospital patients. *Biomed Res Int*, 2014, 931372.
- BLIGHE K, R. S., LEWIS M 2020. *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labelin* [Online]. Available: <https://github.com/kevinblighe/EnhancedVolcano> [Accessed].
- BOCCARD, J. & RUTLEDGE, D. N. 2013. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal Chim Acta*, 769, 30-9.
- BOISVERT, H. & DUNCAN, M. J. 2008. Clathrin-dependent entry of a gingipain adhesin peptide and *Porphyromonas gingivalis* into host cells. *Cell Microbiol*, 10, 2538-52.
- BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-20.
- BOR, B., CEN, L., AGNELLO, M., SHI, W. & HE, X. 2016. Morphological and physiological changes induced by contact-dependent interaction between *Candida albicans* and *Fusobacterium nucleatum*. *Sci Rep*, 6, 27956.
- BÖTTCHER, B., HOFFMANN, B., GARBE, E., WEISE, T., CSERESNYÉS, Z., BRANDT, P., DIETRICH, S., DRIESCH, D., FIGGE, M. T. & VYLKOVA, S. 2020. The Transcription Factor Stp2 Is Important for *Candida albicans* Biofilm Establishment and Sustainability. *Frontiers in Microbiology*, 11, 794.
- BOUHADDANI, S. E., UH, H.-W., JONGBLOED, G., HAYWARD, C., KLARIĆ, L., KIEŁBASA, S. M. & HOUWING-DUISTERMAAT, J. 2018. Integrating omics datasets with the OmicsPLS package. *BMC Bioinformatics*, 19, 371.
- BROWN, A. J., BROWN, G. D., NETEA, M. G. & GOW, N. A. 2014a. Metabolism impacts upon *Candida* immunogenicity and pathogenicity at multiple levels. *Trends Microbiol*, 22, 614-22.
- BROWN, A. J. P., BROWN, G. D., NETEA, M. G. & GOW, N. A. R. 2014b. Metabolism impacts upon *Candida* immunogenicity and pathogenicity at multiple levels. *Trends in Microbiology*, 22, 614-622.
- BRUSCA, M. I., ROSA, A., ALBAINA, O., MORAGUES, M. D., VERDUGO, F. & PONTON, J. 2010. The impact of oral contraceptives on women's periodontal health and the subgingival occurrence of aggressive periodontopathogens and *Candida* species. *J Periodontol*, 81, 1010-8.
- BURGAIN, A., PIC, E., MARKEY, L., TEBBJI, F., KUMAMOTO, C. A. & SELLAM, A. 2019. A novel genetic circuitry governing hypoxic metabolic flexibility, commensalism and virulence in the fungal pathogen *Candida albicans*. *PLoS Pathog*, 15, e1007823.

References

- BURGIN, A., TEBBII, F., KHEMIRI, I. & SELLAM, A. 2020. Metabolic Reprogramming in the Opportunistic Yeast *Candida albicans* in Response to Hypoxia. *mSphere*, 5.
- CALLAHAN, B. J., MCMURDIE, P. J. & HOLMES, S. P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, 11, 2639-2643.
- CALLAHAN, B. J., MCMURDIE, P. J., ROSEN, M. J., HAN, A. W., JOHNSON, A. J. A. & HOLMES, S. P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581-583.
- CANABARRO, A., VALLE, C., FARIAS, M. R., SANTOS, F. B., LAZERA, M. & WANKE, B. 2013. Association of subgingival colonization of *Candida albicans* and other yeasts with severity of chronic periodontitis. *J Periodontal Res*, 48, 428-32.
- CAO, K.-A., BOITARD, S. & BESSE, P. 2011. Sparse PLS Discriminant Analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics*, 12, 253.
- CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., HUNTLEY, J., FIERER, N., OWENS, S. M., BETLEY, J., FRASER, L., BAUER, M., GORMLEY, N., GILBERT, J. A., SMITH, G. & KNIGHT, R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*, 6, 1621-4.
- CAVALCANTI, I. M., NOBBS, A. H., RICOMINI-FILHO, A. P., JENKINSON, H. F. & DEL BEL CURY, A. A. 2016a. Interkingdom cooperation between *Candida albicans*, *Streptococcus oralis* and *Actinomyces oris* modulates early biofilm development on denture material. *Pathog Dis*, 74.
- CAVALCANTI, Y. W., WILSON, M., LEWIS, M., DEL-BEL-CURY, A. A., DA SILVA, W. J. & WILLIAMS, D. W. 2016b. Modulation of *Candida albicans* virulence by bacterial biofilms on titanium surfaces. *Biofouling*, 32, 123-34.
- CAVALHEIRO, M. & TEIXEIRA, M. C. 2018. *Candida* Biofilms: Threats, Challenges, and Promising Strategies. *Frontiers in Medicine*, 5.
- CAVILL, R., JENNEN, D., KLEINJANS, J. & BRIEDE, J. J. 2016. Transcriptomic and metabolomic data integration. *Brief Bioinform*, 17, 891-901.
- CAVILL, R., JENNEN, D., KLEINJANS, J. & BRIEDÉ, J. J. 2015. Transcriptomic and metabolomic data integration. *Briefings in Bioinformatics*, 17, 891-901.
- CHANDRA, J., KUHN, D. M., MUKHERJEE, P. K., HOYER, L. L., MCCORMICK, T. & GHANNOUM, M. A. 2001. Biofilm formation by the fungal pathogen *Candida albicans*: development, architecture, and drug resistance. *J Bacteriol*, 183, 5385-94.
- CHANDRA, J., MCCORMICK, T. S., IMAMURA, Y., MUKHERJEE, P. K. & GHANNOUM, M. A. 2007. Interaction of *Candida albicans* with adherent human peripheral blood mononuclear cells increases *C. albicans* biofilm formation and results in differential expression of pro- and anti-inflammatory cytokines. *Infect Immun*, 75, 2612-20.
- CHEN, C., HEMME, C., BELENO, J., SHI, Z. J., NING, D., QIN, Y., TU, Q., JORGENSEN, M., HE, Z., WU, L. & ZHOU, J. 2018. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. *ISME J*.
- CHEN, C., HUANG, H. & WU, C. H. 2017. Protein Bioinformatics Databases and Resources. *Methods Mol Biol*, 1558, 3-39.
- CHEN, T. & DEWHIRST, F. 2013. Human Oral Microbiome Database (HOMD). In: NELSON, K. E. (ed.) *Encyclopedia of Metagenomics*. New York, NY: Springer New York.

References

- CHENG, Y., YAM, J. K. H., CAI, Z., DING, Y., ZHANG, L. H., DENG, Y. & YANG, L. 2019. Population dynamics and transcriptomic responses of *Pseudomonas aeruginosa* in a complex laboratory microbial community. *NPJ Biofilms Microbiomes*, 5, 1.
- CHO, T., NAGAO, J., IMAYOSHI, R. & TANAKA, Y. 2014. Importance of Diversity in the Oral Microbiota including *Candida* Species Revealed by High-Throughput Technologies. *Int J Dent*, 2014, 454391.
- CHONG, J., SOUFAN, O., LI, C., CARAUS, I., LI, S., BOURQUE, G., WISHART, D. S. & XIA, J. 2018a. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*, 46, W486-W494.
- CHONG, P. P., CHIN, V. K., WONG, W. F., MADHAVAN, P., YONG, V. C. & LOOI, C. Y. 2018b. Transcriptomic and Genomic Approaches for Unravelling *Candida albicans* Biofilm Formation and Drug Resistance-An Update. *Genes (Basel)*, 9.
- COCO, B., BAGG, J., CROSS, L., JOSE, A., CROSS, J. & RAMAGE, G. 2008a. Mixed *Candida albicans* and *Candida glabrata* populations associated with the pathogenesis of denture stomatitis. *Molecular Oral Microbiology*, 23, 377-383.
- COCO, B. J., BAGG, J., CROSS, L. J., JOSE, A., CROSS, J. & RAMAGE, G. 2008b. Mixed *Candida albicans* and *Candida glabrata* populations associated with the pathogenesis of denture stomatitis. *Oral Microbiol Immunol*, 23, 377-83.
- COENYE, T. 2021. Do results obtained with RNA-sequencing require independent verification? *Biofilm*, 3, 100043.
- COLE, J. R., WANG, Q., CARDENAS, E., FISH, J., CHAI, B., FARRIS, R. J., KULAM-SYED-MOHIDEEN, A. S., MCGARRELL, D. M., MARSH, T., GARRITY, G. M. & TIEDJE, J. M. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*, 37, D141-5.
- COLE, J. R., WANG, Q., FISH, J. A., CHAI, B., MCGARRELL, D. M., SUN, Y., BROWN, C. T., PORRAS-ALFARO, A., KUSKE, C. R. & TIEDJE, J. M. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*, 42, D633-42.
- CONESA, A., MADRIGAL, P., TARAZONA, S., GOMEZ-CABRERO, D., CERVERA, A., MCPHERSON, A., SZCZESNIAK, M. W., GAFFNEY, D. J., ELO, L. L., ZHANG, X. & MORTAZAVI, A. 2016a. A survey of best practices for RNA-seq data analysis. *Genome Biol*, 17, 13.
- CONESA, A., MADRIGAL, P., TARAZONA, S., GOMEZ-CABRERO, D., CERVERA, A., MCPHERSON, A., SZCZEŚNIAK, M. W., GAFFNEY, D. J., ELO, L. L., ZHANG, X. & MORTAZAVI, A. 2016b. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17, 13.
- CONWAY, J. R., LEX, A. & GEHLENBORG, N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33, 2938-2940.
- CORNELY, O. A., BASSETTI, M., CALANDRA, T., GARBINO, J., KULLBERG, B. J., LORTHOLARY, O., MEERSSEMAN, W., AKOVA, M., ARENDRUP, M. C., ARIKAN-AKDAGLI, S., BILLE, J., CASTAGNOLA, E., CUENCA-ESTRELLA, M., DONNELLY, J. P., GROLL, A. H., HERBRECHT, R., HOPE, W. W., JENSEN, H. E., LASS-FLORL, C., PETRIKKOS, G., RICHARDSON, M. D., ROILIDES, E., VERWEIJ, P. E., VISCOLI, C., ULLMANN, A. J. & GROUP, E. F. I. S. 2012. ESCMID* guideline for the diagnosis and management of *Candida* diseases 2012: non-neutropenic adult patients. *Clin Microbiol Infect*, 18 Suppl 7, 19-37.
- COTTRET, L., FRAINAY, C., CHAZALVIEL, M., CABANETTES, F., GLOAGUEN, Y., CAMENEN, E., MERLET, B., HEUX, S., PORTAIS, J. C., POUPIN, N., VINSON, F. & JOURDAN, F. 2018.

References

- MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Res*, 46, W495-w502.
- COTTRET, L., WILDRIDGE, D., VINSON, F., BARRETT, M. P., CHARLES, H., SAGOT, M. F. & JOURDAN, F. 2010. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res*, 38, W132-7.
- CRUICKSHANK-QUINN, C. I., JACOBSON, S., HUGHES, G., POWELL, R. L., PETRACHE, I., KECHRIS, K., BOWLER, R. & REISDORPH, N. 2018. Metabolomics and transcriptomics pathway approach reveals outcome-specific perturbations in COPD. *Scientific Reports*, 8, 17132.
- CUI, L., MORRIS, A. & GHEDIN, E. 2013. The human mycobiome in health and disease. *Genome Med*, 5, 63.
- CULLEN, C. M., ANEJA, K. K., BEYHAN, S., CHO, C. E., WOLOSZYNEK, S., CONVERTINO, M., MCCOY, S. J., ZHANG, Y., ANDERSON, M. Z., ALVAREZ-PONCE, D., SMIRNOVA, E., KARSTENS, L., DORRESTEIN, P. C., LI, H., SEN GUPTA, A., CHEUNG, K., POWERS, J. G., ZHAO, Z. & ROSEN, G. L. 2020. Emerging Priorities for Microbiome Research. *Front Microbiol*, 11, 136.
- DABROWA, N., TAXER, S. S. & HOWARD, D. H. 1976. Germination of *Candida albicans* induced by proline. *Infect Immun*, 13, 830-5.
- DANHOF, H. A. & LORENZ, M. C. 2015. The *Candida albicans* ATO Gene Family Promotes Neutralization of the Macrophage Phagolysosome. *Infect Immun*, 83, 4416-26.
- DE GROOT, P. W. J., DE BOER, A. D., CUNNINGHAM, J., DEKKER, H. L., DE JONG, L., HELLINGWERF, K. J., DE KOSTER, C. & KLIS, F. M. 2004. Proteomic Analysis of *Candida albicans* Cell Walls Reveals Covalently Bound Carbohydrate-Active Enzymes and Adhesins. *Eukaryotic Cell*, 3, 955-965.
- DE PAIVA, C. S., JONES, D. B., STERN, M. E., BIAN, F., MOORE, Q. L., CORBIERE, S., STRECKFUS, C. F., HUTCHINSON, D. S., AJAMI, N. J., PETROSINO, J. F. & PFLUGFELDER, S. C. 2016. Altered Mucosal Microbiome Diversity and Disease Severity in Sjogren Syndrome. *Sci Rep*, 6, 23561.
- DELANEY, C., KEAN, R., SHORT, B., TUMELTY, M., NILE, C. J., MCLEAN, W. & RAMAGE, G. 2018. Fungi at the scene of the crime: innocent bystanders or accomplices in oral infections? *Current Clinical Microbiology Reports*, 5, 190-200.
- DESANTIS, T. Z., HUGENHOLTZ, P., LARSEN, N., ROJAS, M., BRODIE, E. L., KELLER, K., HUBER, T., DALEVI, D., HU, P. & ANDERSEN, G. L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*, 72, 5069-72.
- DEWHIRST, F. E., CHEN, T., IZARD, J., PASTER, B. J., TANNER, A. C., YU, W. H., LAKSHMANAN, A. & WADE, W. G. 2010. The human oral microbiome. *J Bacteriol*, 192, 5002-17.
- DIAZ, P. I., XIE, Z., SOBUE, T., THOMPSON, A., BIYIKOGLU, B., RICKER, A., IKONOMOU, L. & DONGARI-BAGTZOGLOU, A. 2012. Synergistic interaction between *Candida albicans* and commensal oral streptococci in a novel in vitro mucosal model. *Infect Immun*, 80, 620-32.
- DIKBAS, I., KOKSAL, T. & CALIKKOCAOGLU, S. 2006. Investigation of the cleanliness of dentures in a university hospital. *Int J Prosthodont*, 19, 294-8.
- DOLLIVE, S., PETERFREUND, G. L., SHERRILL-MIX, S., BITTINGER, K., SINHA, R., HOFFMANN, C., NABEL, C. S., HILL, D. A., ARTIS, D., BACHMAN, M. A., CUSTERS-ALLEN, R., GRUNBERG, S., WU, G. D., LEWIS, J. D. & BUSHMAN, F. D. 2012. A tool kit for

References

- quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol*, 13, R60.
- DUDLEY, E. & BOND, A. E. 2014. Chapter Two - Phosphoproteomic Techniques and Applications. In: DONEV, R. (ed.) *Advances in Protein Chemistry and Structural Biology*. Academic Press.
- DUPUY, A. K., DAVID, M. S., LI, L., HEIDER, T. N., PETERSON, J. D., MONTANO, E. A., DONGARI-BAGTZOGLOU, A., DIAZ, P. I. & STRAUSBAUGH, L. D. 2014. Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of *Malassezia* as a prominent commensal. *PLoS One*, 9, e90899.
- DUTTON, L. C., NOBBS, A. H., JEPSON, K., JEPSON, M. A., VICKERMAN, M. M., AQEEL ALAWFI, S., MUNRO, C. A., LAMONT, R. J. & JENKINSON, H. F. 2014. O-mannosylation in *Candida albicans* enables development of interkingdom biofilm communities. *mBio*, 5, e00911.
- DUTTON, L. C., PASZKIEWICZ, K. H., SILVERMAN, R. J., SPLATT, P. R., SHAW, S., NOBBS, A. H., LAMONT, R. J., JENKINSON, H. F. & RAMSDALE, M. 2016. Transcriptional landscape of trans-kingdom communication between *Candida albicans* and *Streptococcus gordonii*. *Mol Oral Microbiol*, 31, 136-61.
- EDGAR, R. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv.
- EDGAR, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*, 10, 996-8.
- EDLUND, A., YANG, Y., YOOSEPH, S., HE, X., SHI, W. & MCLEAN, J. S. 2018. Uncovering complex microbiome activities via metatranscriptomics during 24 hours of oral biofilm assembly and maturation. *Microbiome*, 6, 217.
- EKE, P. I., DYE, B. A., WEI, L., SLADE, G. D., THORNTON-EVANS, G. O., BORGNACKE, W. S., TAYLOR, G. W., PAGE, R. C., BECK, J. D. & GENCO, R. J. 2015. Update on Prevalence of Periodontitis in Adults in the United States: NHANES 2009 to 2012. *J Periodontol*, 86, 611-22.
- ELLEPOLA, K., TRUONG, T., LIU, Y., LIN, Q., LIM, T. K., LEE, Y. M., CAO, T., KOO, H. & SENEVIRATNE, C. J. 2019. Multi-omics Analyses Reveal Synergistic Carbohydrate Metabolism in *Streptococcus mutans*-*Candida albicans* Mixed-Species Biofilms. *Infect Immun*, 87.
- ELLS, R., KOCK, J. L. F., VAN WYK, P. W. J., BOTES, P. J. & POHL, C. H. 2008. Arachidonic acid increases antifungal susceptibility of *Candida albicans* and *Candida dubliniensis*. *Journal of Antimicrobial Chemotherapy*, 63, 124-128.
- ENE, I. V., WALKER, L. A., SCHIAVONE, M., LEE, K. K., MARTIN-YKEN, H., DAGUE, E., GOW, N. A., MUNRO, C. A. & BROWN, A. J. 2015. Cell Wall Remodeling Enzymes Modulate Fungal Cell Wall Elasticity and Osmotic Stress Resistance. *MBio*, 6, e00986.
- ERLANDSEN, S. L., KRISTICH, C. J., DUNNY, G. M. & WELLS, C. L. 2004. High-resolution visualization of the microbial glycocalyx with low-voltage scanning electron microscopy: dependence on cationic dyes. *J Histochem Cytochem*, 52, 1427-35.
- EVERAERT, C., LUYPAERT, M., MAAG, J. L. V., CHENG, Q. X., DINGER, M. E., HELLEMANS, J. & MESTDAGH, P. 2017. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci Rep*, 7, 1559.
- FARRER, R. A. & FISHER, M. C. 2017. Chapter Three - Describing Genomic and Epigenomic Traits Underpinning Emerging Fungal Pathogens. In: TOWNSEND, J. P. & WANG, Z. (eds.) *Advances in Genetics*. Academic Press.

References

- FENG, Q., SUMMERS, E., GUO, B. & FINK, G. 1999. Ras signaling is required for serum-induced hyphal differentiation in *Candida albicans*. *J Bacteriol*, 181, 6339-46.
- FINKEL, J. S. & MITCHELL, A. P. 2011. Genetic control of *Candida albicans* biofilm development. *Nat Rev Microbiol*, 9, 109-18.
- FOX, E. P., BUI, C. K., NETT, J. E., HARTOONI, N., MUI, M. C., ANDES, D. R., NOBILE, C. J. & JOHNSON, A. D. 2015. An expanded regulatory network temporally controls *Candida albicans* biofilm formation. *Mol Microbiol*, 96, 1226-39.
- FOX, E. P., COWLEY, E. S., NOBILE, C. J., HARTOONI, N., NEWMAN, D. K. & JOHNSON, A. D. 2014. Anaerobic bacteria grow within *Candida albicans* biofilms and induce biofilm formation in suspension cultures. *Curr Biol*, 24, 2411-6.
- FUKUSHIMA, A., KANAYA, S. & NISHIDA, K. 2014. Integrated network analysis and effective tools in plant systems biology. *Front Plant Sci*, 5, 598.
- FUX, C. A., SHIRTLIFF, M., STOODLEY, P. & COSTERTON, J. W. 2005. Can laboratory reference strains mirror "real-world" pathogenesis? *Trends Microbiol*, 13, 58-63.
- GALIMANAS, V., HALL, M. W., SINGH, N., LYNCH, M. D., GOLDBERG, M., TENENBAUM, H., CVITKOVITCH, D. G., NEUFELD, J. D. & SENADHEERA, D. B. 2014. Bacterial community composition of chronic periodontitis and novel oral sampling sites for detecting disease indicators. *Microbiome*, 2, 32.
- GANGULY, S. & MITCHELL, A. P. 2011. Mucosal biofilms of *Candida albicans*. *Curr Opin Microbiol*, 14, 380-5.
- GARBE, E. & VYLKOVA, S. 2019. Role of Amino Acid Metabolism in the Virulence of Human Pathogenic Fungi. *Current Clinical Microbiology Reports*, 6.
- GENDREAU, L. & LOEWY, Z. G. 2011. Epidemiology and etiology of denture stomatitis. *J Prosthodont*, 20, 251-60.
- GHANNOUM, M. A., JUREVIC, R. J., MUKHERJEE, P. K., CUI, F., SIKAROODI, M., NAQVI, A. & GILLEVET, P. M. 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog*, 6, e1000713.
- GHORAI, P., IRFAN, M., NARULA, A. & DATTA, A. 2018. A comprehensive analysis of *Candida albicans* phosphoproteome reveals dynamic changes in phosphoprotein abundance during hyphal morphogenesis. *Applied Microbiology and Biotechnology*, 102, 9731-9743.
- GHOSH, S., NAVARATHNA, D., ROBERTS, D., COOPER, J., ATKIN, A., PETRO, T. & NICKERSON, K. 2009. Arginine-Induced Germ Tube Formation in *Candida albicans* Is Essential for Escape from Murine Macrophage Line RAW 264.7. *Infection and immunity*, 77, 1596-605.
- GILOTEAUX, L., GOODRICH, J. K., WALTERS, W. A., LEVINE, S. M., LEY, R. E. & HANSON, M. R. 2016. Reduced diversity and altered composition of the gut microbiome in individuals with myalgic encephalomyelitis/chronic fatigue syndrome. *Microbiome*, 4, 30.
- GLOAGUEN, Y., MORTON, F., DALY, R., GURDEN, R., ROGERS, S., WANDY, J., WILSON, D., BARRETT, M. & BURGESS, K. 2017. PiMP my metabolome: an integrated, web-based tool for LC-MS metabolomics data. *Bioinformatics (Oxford, England)*, 33, 4007-4009.
- GONG, Y., LI, T., YU, C. & SUN, S. 2017. *Candida albicans* Heat Shock Proteins and Hsps-Associated Signaling Pathways as Potential Antifungal Targets. *Front Cell Infect Microbiol*, 7, 520.
- GONZÁLEZ, I., CAO, K. A., DAVIS, M. J. & DÉJEAN, S. 2012. Visualising associations between paired 'omics' data sets. *BioData Min*, 5, 19.

References

- GOUDELA, S. & TSILIVI, H. 2006. Comparative kinetic analysis of AzgA and Fcy21p, prototypes of the two major fungal hypoxanthine-adenine-guanine transporter families. *Molecular membrane biology*, 23, 291-303.
- GRIFFEN, A. L., BEALL, C. J., CAMPBELL, J. H., FIRESTONE, N. D., KUMAR, P. S., YANG, Z. K., PODAR, M. & LEYS, E. J. 2012. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J*, 6, 1176-85.
- GRIFFIN, J. L., BONNEY, S. A., MANN, C., HEBBACHI, A. M., GIBBONS, G. F., NICHOLSON, J. K., SHOULDERS, C. C. & SCOTT, J. 2004. An integrated reverse functional genomic and metabolic approach to understanding orotic acid-induced fatty liver. *Physiol Genomics*, 17, 140-9.
- HAFFAJEE, A. D., SOCRANSKY, S. S., PATEL, M. R. & SONG, X. 2008. Microbial complexes in supragingival plaque. *Oral Microbiol Immunol*, 23, 196-205.
- HAJISHENGALLIS, G., DARVEAU, R. P. & CURTIS, M. A. 2012. The keystone-pathogen hypothesis. *Nat Rev Microbiol*, 10, 717-25.
- HAJISHENGALLIS, G. & LAMONT, R. J. 2016. Dancing with the Stars: How Choreographed Bacterial Interactions Dictate Nososymbiocity and Give Rise to Keystone Pathogens, Accessory Pathogens, and Pathobionts. *Trends Microbiol*, 24, 477-489.
- HAN, T.-L., CANNON, R. & VILLAS-BÔAS, S. 2012. Metabolome analysis during the morphological transition of *Candida albicans*. *Metabolomics*, 8, 1204-1217.
- HAN, T.-L., CANNON, R. D., GALLO, S. M. & VILLAS-BÔAS, S. G. 2019a. A metabolomic study of the effect of *Candida albicans* glutamate dehydrogenase deletion on growth and morphogenesis. *npj Biofilms and Microbiomes*, 5, 13.
- HAN, T. L., CANNON, R. D., GALLO, S. M. & VILLAS-BÔAS, S. G. 2019b. A metabolomic study of the effect of *Candida albicans* glutamate dehydrogenase deletion on growth and morphogenesis. *NPJ Biofilms Microbiomes*, 5, 13.
- HANNAH, V. E., O'DONNELL, L., ROBERTSON, D. & RAMAGE, G. 2017. Denture Stomatitis: Causes, Cures and Prevention. *Prim Dent J*, 6, 46-51.
- HARRIOTT, M. M. & NOVERR, M. C. 2011. Importance of *Candida*-bacterial polymicrobial biofilms in disease. *Trends Microbiol*, 19, 557-63.
- HASAN, F., XESS, I., WANG, X., JAIN, N. & FRIES, B. C. 2009. Biofilm formation in clinical *Candida* isolates and its association with virulence. *Microbes Infect*, 11, 753-61.
- HAWSER, S. 1996. Comparisons of the susceptibilities of planktonic and adherent *Candida albicans* to antifungal agents: a modified XTT tetrazolium assay using synchronised *C. albicans* cells. *J Med Vet Mycol*, 34, 149-52.
- HAWSER, S. P., NORRIS, H., JESSUP, C. J. & GHANNOUM, M. A. 1998. Comparison of a 2,3-bis(2-methoxy-4-nitro-5-sulphophenyl)-5-[(phenylamino)carbonyl]-2H-tetrazolium hydroxide (XTT) colorimetric method with the standardized National Committee for Clinical Laboratory Standards method of testing clinical yeast isolates for susceptibility to antifungal agents. *J Clin Microbiol*, 36, 1450-2.
- HE, J., KIM, D., ZHOU, X., AHN, S.-J., BURNE, R. A., RICHARDS, V. P. & KOO, H. 2017. RNA-Seq Reveals Enhanced Sugar Metabolism in *Streptococcus mutans* Co-cultured with *Candida albicans* within Mixed-Species Biofilms. *Frontiers in Microbiology*, 8, 1036.
- HERNÁNDEZ-DE-DIEGO, R., TARAZONA, S., MARTÍNEZ-MIRA, C., BALZANO-NOGUEIRA, L., FURIÓ-TARÍ, P., PAPPAS, G. J., JR. & CONESA, A. 2018. PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Res*, 46, W503-w509.

References

- HIRSCHFELD, J., WHITE, P. C., MILWARD, M. R., COOPER, P. R. & CHAPPLE, I. L. C. 2017. Modulation of Neutrophil Extracellular Trap and Reactive Oxygen Species Release by Periodontal Bacteria. *Infect Immun*, 85.
- HOARAU, G., MUKHERJEE, P. K., GOWER-ROUSSEAU, C., HAGER, C., CHANDRA, J., RETUERTO, M. A., NEUT, C., VERMEIRE, S., CLEMENTE, J., COLOMBEL, J. F., FUJIOKA, H., POULAIN, D., SENDID, B. & GHANNOUM, M. A. 2016. Bacteriome and Mycobiome Interactions Underscore Microbial Dysbiosis in Familial Crohn's Disease. *MBio*, 7.
- HOLMES, A. R. & SHEPHERD, M. G. 1987. Proline-induced germ-tube formation in *Candida albicans*: role of proline uptake and nitrogen metabolism. *J Gen Microbiol*, 133, 3219-28.
- HORGAN, R. P. & KENNY, L. C. 2011. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, 13, 189-195.
- HU, Q., NOLL, R. J., LI, H., MAKAROV, A., HARDMAN, M. & GRAHAM COOKS, R. 2005. The Orbitrap: a new mass spectrometer. *J Mass Spectrom*, 40, 430-43.
- HUFFNAGLE, G. B. & NOVERR, M. C. 2013. The emerging world of the fungal microbiome. *Trends Microbiol*, 21, 334-41.
- HUMAN MICROBIOME PROJECT, C. 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, 486, 207-14.
- IBRAHIM-GRANET, O., DUBOURDEAU, M., LATGÉ, J. P., AVE, P., HUERRE, M., BRAKHAGE, A. & BROCK, M. 2008. Methylcitrate synthase from *Aspergillus fumigatus* is essential for manifestation of invasive aspergillosis. *Cell Microbiol*, 10, 134-48.
- JACOBSEN, M. D., BEYNON, R. J., GETHINGS, L. A., CLAYDON, A. J., LANGRIDGE, J. I., VISSERS, J. P. C., BROWN, A. J. P. & HAMMOND, D. E. 2018. Specificity of the osmotic stress response in *Candida albicans* highlighted by quantitative proteomics. *Scientific Reports*, 8, 14492.
- JAIN, N., KOHLI, R., COOK, E., GIALANELLA, P., CHANG, T. & FRIES, B. C. 2007. Biofilm formation by and antifungal susceptibility of *Candida* isolates from urine. *Appl Environ Microbiol*, 73, 1697-703.
- JANUS, M. M., WILLEMS, H. M. & KROM, B. P. 2016. *Candida albicans* in Multispecies Oral Communities; A Keystone Commensal? *Adv Exp Med Biol*, 931, 13-20.
- JASSAL, B., MATTHEWS, L., VITERI, G., GONG, C., LORENTE, P., FABREGAT, A., SIDIROPOULOS, K., COOK, J., GILLESPIE, M., HAW, R., LONEY, F., MAY, B., MILACIC, M., ROTHFELS, K., SEVILLA, C., SHAMOVSKY, V., SHORSER, S., VARUSAI, T., WEISER, J., WU, G., STEIN, L., HERMJAKOB, H. & D'EUSTACHIO, P. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res*, 48, D498-d503.
- JASTRZĘBOWSKA, K. & GABRIEL, I. 2015. Inhibitors of amino acids biosynthesis as antifungal agents. *Amino Acids*, 47, 227-49.
- JENKINSON, H. F. 2011. Beyond the oral microbiome. *Environ Microbiol*, 13, 3077-87.
- JOO, M. Y., SHIN, J. H., JANG, H. C., SONG, E. S., KEE, S. J., SHIN, M. G., SUH, S. P. & RYANG, D. W. 2013. Expression of SAP5 and SAP9 in *Candida albicans* biofilms: comparison of bloodstream isolates with isolates from other sources. *Med Mycol*, 51, 892-6.
- JUNG, W. H. 2015. The Zinc Transport Systems and Their Regulation in Pathogenic Fungi. *Mycobiology*, 43, 179-83.
- KA, L. C., GONZÁLEZ, I. & DÉJEAN, S. 2009. integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics*, 25, 2855-6.
- KA, L. C., ROSSOUW, D., ROBERT-GRANIÉ, C. & BESSE, P. 2008. A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol*, 7, Article 35.

References

- KAMBUROV, A., CAVILL, R., EBBELS, T. M. D., HERWIG, R. & KEUN, H. C. 2011. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics*, 27, 2917-2918.
- KANEHISA, M., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44, D457-D462.
- KARPIEVITCH, Y. V., POLPITIYA, A. D., ANDERSON, G. A., SMITH, R. D. & DABNEY, A. R. 2010. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *The annals of applied statistics*, 4, 1797-1823.
- KAUR, S. & MISHRA, P. 1991. Amino acid uptake as a function of differentiation in *Candida albicans*: studies of a non-germinative variant. *FEMS Microbiol Lett*, 66, 341-4.
- KEAN, R., DELANEY, C., RAJENDRAN, R., SHERRY, L., METCALFE, R., THOMAS, R., MCLEAN, W., WILLIAMS, C. & RAMAGE, G. 2018a. Gaining Insights from *Candida* Biofilm Heterogeneity: One Size Does Not Fit All. *J Fungi (Basel)*, 4.
- KEAN, R., DELANEY, C., SHERRY, L., BORMAN, A., JOHNSON, E. M., RICHARDSON, M. D., RAUTEMAA-RICHARDSON, R., WILLIAMS, C. & RAMAGE, G. 2018b. Transcriptome Assembly and Profiling of *Candida auris* Reveals Novel Insights into Biofilm-Mediated Resistance. *mSphere*, 3.
- KEAN, R., RAJENDRAN, R., HAGGARTY, J., TOWNSEND, E. M., SHORT, B., BURGESS, K. E., LANG, S., MILLINGTON, O., MACKAY, W. G., WILLIAMS, C. & RAMAGE, G. 2017. *Candida albicans* mycofilms support *Staphylococcus aureus* colonization and enhances miconazole resistance in dual-species interactions. *Front Microbiol*, 8, 258.
- KELL, D. B., BROWN, M., DAVEY, H. M., DUNN, W. B., SPASIC, I. & OLIVER, S. G. 2005. Metabolic footprinting and systems biology: the medium is the message. *Nature Reviews Microbiology*, 3, 557-565.
- KHO, Z. Y. & LAL, S. K. 2018. The Human Gut Microbiome – A Potential Controller of Wellness and Disease. *Frontiers in Microbiology*, 9.
- KIM, D., PAGGI, J. M., PARK, C., BENNETT, C. & SALZBERG, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, 37, 907-915.
- KLASSON, H., FINK, G. R. & LJUNGDAHL, P. O. 1999. Ssy1p and Ptr3p are plasma membrane components of a yeast system that senses extracellular amino acids. *Mol Cell Biol*, 19, 5405-16.
- KLUG, L. & DAUM, G. 2014. Yeast lipid metabolism at a glance. *FEMS Yeast Research*, 14, 369-388.
- KOEHLER, P., TACKE, D. & CORNELLY, O. A. 2014. Our 2014 approach to candidaemia. *Mycoses*, 57, 581-3.
- KOJIC, E. M. & DAROUICHE, R. O. 2004. *Candida* infections of medical devices. *Clin Microbiol Rev*, 17, 255-67.
- KOZICH, J. J., WESTCOTT, S. L., BAXTER, N. T., HIGHLANDER, S. K. & SCHLOSS, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*, 79, 5112-20.
- KRAIDLOVA, L., SCHREVEENS, S., TOURNU, H., VAN ZEEBROECK, G., SYCHROVA, H. & VAN DIJCK, P. 2016. Characterization of the *Candida albicans* Amino Acid Permease Family: Gap2 Is the Only General Amino Acid Permease and Gap4 Is an S-

References

- Adenosylmethionine (SAM) Transporter Required for SAM-Induced Morphogenesis. *mSphere*, 1.
- KRANEVELD, E. A., BUIJS, M. J., BONDER, M. J., VISSER, M., KEIJSER, B. J., CRIELAARD, W. & ZAURA, E. 2012. The relation between oral Candida load and bacterial microbiome profiles in Dutch older adults. *PLoS One*, 7, e42770.
- KUCHARIKOVA, S., TOURNU, H., HOLTAPPELS, M., VAN DIJCK, P. & LAGROU, K. 2010. In vivo efficacy of anidulafungin against mature Candida albicans biofilms in a novel rat model of catheter-associated Candidiasis. *Antimicrob Agents Chemother*, 54, 4474-5.
- KUCHARÍKOVÁ, S., TOURNU, H., LAGROU, K., VAN DIJCK, P. & BUJDÁKOVÁ, H. 2011. Detailed comparison of Candida albicans and Candida glabrata biofilms under different conditions and their susceptibility to caspofungin and anidulafungin. *J Med Microbiol*, 60, 1261-1269.
- KUHN, D. M., BALKIS, M., CHANDRA, J., MUKHERJEE, P. K. & GHANNOUM, M. A. 2003. Uses and limitations of the XTT assay in studies of Candida growth and metabolism. *J Clin Microbiol*, 41, 506-8.
- KUMAR, C. P. & MENON, T. 2006. Biofilm production by clinical isolates of Candida species. *Med Mycol*, 44, 99-101.
- KURAKADO, S., ARAI, R. & SUGITA, T. 2018. Association of the hypha-related protein Pra1 and zinc transporter Zrt1 with biofilm formation by the pathogenic yeast Candida albicans. *Microbiol Immunol*, 62, 405-410.
- LAMBOOIJ, J. M., HOOGENKAMP, M. A., BRANDT, B. W., JANUS, M. M. & KROM, B. P. 2017. Fungal mitochondrial oxygen consumption induces the growth of strict anaerobic bacteria. *Fungal Genet Biol*, 109, 1-6.
- LANGILLE, M. G., ZANEVELD, J., CAPORASO, J. G., MCDONALD, D., KNIGHTS, D., REYES, J. A., CLEMENTE, J. C., BURKEPILE, D. E., VEGA THURBER, R. L., KNIGHT, R., BEIKO, R. G. & HUTTENHOWER, C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*, 31, 814-21.
- LEDERBERG, J. & MCCRAY, A. 2001. The Scientist : 'Ome Sweet 'Omics-- A Genealogical Treasury of Words. *The Scientist*, 17.
- LEE, I. R., LUI, E. Y., CHOW, E. W., ARRAS, S. D., MORROW, C. A. & FRASER, J. A. 2013. Reactive oxygen species homeostasis and virulence of the fungal pathogen Cryptococcus neoformans requires an intact proline catabolism pathway. *Genetics*, 194, 421-33.
- LI, F., SVAROVSKY, M. J., KARLSSON, A. J., WAGNER, J. P., MARCHILLO, K., OSHEL, P., ANDES, D. & PALECEK, S. P. 2007. Eap1p, an adhesin that mediates Candida albicans biofilm formation in vitro and in vivo. *Eukaryot Cell*, 6, 931-9.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNEL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & SUBGROUP, G. P. D. P. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-2079.
- LI, H., LIU, L., ZHANG, S., CUI, W. & LV, J. 2012. Identification of Antifungal Compounds Produced by Lactobacillus casei AST18. *Current Microbiology*, 65, 156-161.
- LI, Y., WANG, K., ZHANG, B., TU, Q., YAO, Y., CUI, B., REN, B., HE, J., SHEN, X., VAN NOSTRAND, J. D., ZHOU, J., SHI, W., XIAO, L., LU, C. & ZHOU, X. 2019. Salivary mycobiome dysbiosis and its potential impact on bacteriome shifts and host immunity in oral lichen planus. *International Journal of Oral Science*, 11, 13.
- LI, Z., CHEN, Y., LIU, D., ZHAO, N., CHENG, H., REN, H., GUO, T., NIU, H., ZHUANG, W., WU, J. & YING, H. 2015. Involvement of glycolysis/gluconeogenesis and signaling regulatory

References

- pathways in *Saccharomyces cerevisiae* biofilms during fermentation. *Front Microbiol*, 6, 139.
- LIU, X., WANG, D., YU, C., LI, T., LIU, J. & SUN, S. 2016. Potential Antifungal Targets against a *Candida* Biofilm Based on an Enzyme in the Arachidonic Acid Cascade-A Review. *Front Microbiol*, 7, 1925.
- LIU, X. L., MING, Y. N., ZHANG, J. Y., CHEN, X. Y., ZENG, M. D. & MAO, Y. M. 2017. Gene-metabolite network analysis in different nonalcoholic fatty liver disease phenotypes. *Exp Mol Med*, 49, e283.
- ŁOBODA, D. & ROWIŃSKA-ŻYREK, M. 2018. *Candida albicans* zincophore and zinc transporter interactions with Zn(ii) and Ni(ii). *Dalton Trans*, 47, 2646-2654.
- LOESCHE, W. J. 1976. Chemotherapy of dental plaque infections. *Oral Sci Rev*, 9, 65-107.
- LÓPEZ-RIBOT, J. L. 2005. *Candida albicans* Biofilms: More Than Filamentation. *Current Biology*, 15, R453-R455.
- LÓPEZ-RIBOT, J. L., ALLOUSH, H. M., MASTEN, B. J. & CHAFFIN, W. L. 1996. Evidence for presence in the cell wall of *Candida albicans* of a protein related to the hsp70 family. *Infect Immun*, 64, 3333-40.
- LÓPEZ-RIBOT, J. L. & CHAFFIN, W. L. 1996. Members of the Hsp70 family of proteins in the cell wall of *Saccharomyces cerevisiae*. *J Bacteriol*, 178, 4724-6.
- LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15, 550.
- LUO, W., PANT, G., BHAVNASI, Y. K., BLANCHARD, S. G., JR. & BROUWER, C. 2017. Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Res*.
- MAGALHÃES, R. S., DE LIMA, K. C., DE ALMEIDA, D. S., DE MESQUITA, J. F. & ELEUTHERIO, E. C. 2017. Trehalose-6-Phosphate as a Potential Lead Candidate for the Development of Tps1 Inhibitors: Insights from the Trehalose Biosynthesis Pathway in Diverse Yeast Species. *Appl Biochem Biotechnol*, 181, 914-924.
- MANTZOURANI, M., GILBERT, S. C., FENLON, M. & BEIGHTON, D. 2010. Non-oral bifidobacteria and the aciduric microbiota of the denture plaque biofilm. *Mol Oral Microbiol*, 25, 190-9.
- MARCOS-ZAMBRANO, L. J., ESCRIBANO, P., BOUZA, E. & GUINEA, J. 2016. Susceptibility of *Candida albicans* biofilms to caspofungin and anidulafungin is not affected by metabolic activity or biomass production. *Med Mycol*, 54, 155-61.
- MARINI, F. & BINDER, H. 2019. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics*, 20, 331.
- MARKLEY, J. L., BRUSCHWEILER, R., EDISON, A. S., EGHBALNIA, H. R., POWERS, R., RAFTERY, D. & WISHART, D. S. 2016. The future of NMR-based metabolomics. *Curr Opin Biotechnol*, 43, 34-40.
- MARSH, P. D. 1994. Microbial ecology of dental plaque and its significance in health and disease. *Adv Dent Res*, 8, 263-71.
- MARSH, P. D. & ZAURA, E. 2017. Dental biofilm: ecological interactions in health and disease. *J Clin Periodontol*, 44 Suppl 18, S12-S22.
- MARTÍNEZ, P. & LJUNGDAHL, P. O. 2005. Divergence of Stp1 and Stp2 transcription factors in *Candida albicans* places virulence factors required for proper nutrient acquisition under amino acid control. *Mol Cell Biol*, 25, 9435-46.
- MARTINEZ, R. C. R., SENEY, S. L., SUMMERS, K. L., NOMIZO, A., DE MARTINIS, E. C. P. & REID, G. 2009. Effect of *Lactobacillus rhamnosus* GR-1 and *Lactobacillus reuteri* RC-14 on

References

- the ability of *Candida albicans* to infect cells and induce inflammation. *Microbiology and Immunology*, 53, 487-495.
- MARTINS, M., HENRIQUES, M., LOPEZ-RIBOT, J. L. & OLIVEIRA, R. 2012. Addition of DNase improves the in vitro activity of antifungal drugs against *Candida albicans* biofilms. *Mycoses*, 55, 80-5.
- MARTINS, M., UPPULURI, P., THOMAS, D. P., CLEARY, I. A., HENRIQUES, M., LOPEZ-RIBOT, J. L. & OLIVEIRA, R. 2010. Presence of extracellular DNA in the *Candida albicans* biofilm matrix and its contribution to biofilms. *Mycopathologia*, 169, 323-31.
- MARTORI, E., AYUSO-MONTERO, R., MARTINEZ-GOMIS, J., VIÑAS, M. & PERAIRE, M. 2014. Risk factors for denture-related oral mucosal lesions in a geriatric population. *Journal of Prosthetic Dentistry*, 111, 273-279.
- MATSUBARA, V. H., WANG, Y., BANDARA, H. M. H. N., MAYER, M. P. A. & SAMARANAYAKE, L. P. 2016. Probiotic lactobacilli inhibit early stages of *Candida albicans* biofilm development by reducing their growth, cell adhesion, and filamentation. *Applied Microbiology and Biotechnology*, 100, 6415-6426.
- MAYER, F. L., WILSON, D. & HUBE, B. 2013. *Candida albicans* pathogenicity mechanisms. *Virulence*, 4, 119-28.
- MCCALL, A. D., PATHIRANA, R. U., PRABHAKAR, A., CULLEN, P. J. & EDGERTON, M. 2019. *Candida albicans* biofilm development is governed by cooperative attachment and adhesion maintenance proteins. *npj Biofilms and Microbiomes*, 5, 21.
- MCCARTHY, M. & WALSH, T. 2018. Amino Acid Metabolism and Transport Mechanisms as Potential Antifungal Targets. *International Journal of Molecular Sciences*, 19, 909.
- MCMURDIE, P. J. & HOLMES, S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8, e61217.
- MCTAGGART, L. R., COPELAND, J. K., SURENDRA, A., WANG, P. W., HUSAIN, S., COBURN, B., GUTTMAN, D. S. & KUS, J. V. 2019. Mycobiome Sequencing and Analysis Applied to Fungal Community Profiling of the Lower Respiratory Tract During Fungal Pathogenesis. *Front Microbiol*, 10, 512.
- MEGSON, Z. A., KOERDT, A., SCHUSTER, H., LUDWIG, R., JANESCH, B., FREY, A., NAYLOR, K., WILSON, I. B., STAFFORD, G. P., MESSNER, P. & SCHAFFER, C. 2015. Characterization of an alpha-L-fucosidase from the periodontal pathogen *Tannerella forsythia*. *Virulence*, 6, 282-92.
- MERMEL, L. A., ALLON, M., BOUZA, E., CRAVEN, D. E., FLYNN, P., O'GRADY, N. P., RAAD, II, RIJNDERS, B. J., SHERERTZ, R. J. & WARREN, D. K. 2009. Clinical practice guidelines for the diagnosis and management of intravascular catheter-related infection: 2009 Update by the Infectious Diseases Society of America. *Clin Infect Dis*, 49, 1-45.
- MESNAGE, R., BISERNI, M., BALU, S., FRAINAY, C., POUPIN, N., JOURDAN, F., WOZNIK, E., XENAKIS, T., MEIN, C. A. & ANTONIOU, M. N. 2018. Integrated transcriptomics and metabolomics reveal signatures of lipid metabolism dysregulation in HepaRG liver cells exposed to PCB 126. *Arch Toxicol*, 92, 2533-2547.
- METWALY, A., DUNKEL, A., WALDSCHMITT, N., RAJ, A. C. D., LAGKOUVARDOS, I., CORRALIZA, A. M., MAYORGAS, A., MARTINEZ-MEDINA, M., REITER, S., SCHLOTER, M., HOFMANN, T., ALLEZ, M., PANES, J., SALAS, A. & HALLER, D. 2020. Integrated microbiota and metabolite profiles link Crohn's disease to sulfur metabolism. *Nature Communications*, 11, 4322.
- MILLHOUSE, E., JOSE, A., SHERRY, L., LAPPIN, D. F., PATEL, N., MIDDLETON, A. M., PRATTEN, J., CULSHAW, S. & RAMAGE, G. 2014. Development of an in vitro periodontal biofilm

References

- model for assessing antimicrobial and host modulatory effects of bioactive molecules. *BMC Oral Health*, 14, 80.
- MIO, T., ADACHI-SHIMIZU, M., TACHIBANA, Y., TABUCHI, H., INOUE, S., YABE, T., YAMADA-OKABE, T., ARISAWA, M., WATANABE, T. & YAMADA-OKABE, H. 1997. Cloning of the *Candida albicans* homolog of *Saccharomyces cerevisiae* GSC1/FKS1 and its involvement in β -1,3-glucan synthesis. *Journal of bacteriology*, 179, 4096-105.
- MIRAMÓN, P. & LORENZ, M. 2016. The SPS amino acid sensor mediates nutrient acquisition and immune evasion in *Candida albicans*. *Cellular microbiology*, 18.
- MISHRA, N. N., ALI, S. & SHUKLA, P. K. 2014. Arachidonic acid affects biofilm formation and PGE2 level in *Candida albicans* and non-*albicans* species in presence of subinhibitory concentration of fluconazole and terbinafine. *Braz J Infect Dis*, 18, 287-93.
- MITCHELL, K. F., ZARNOWSKI, R. & ANDES, D. R. 2016. Fungal Super Glue: The Biofilm Matrix and Its Composition, Assembly, and Functions. *PLoS Pathog*, 12, e1005828.
- MORA-MONTES, H. M., BATES, S., NETEA, M. G., DIAZ-JIMENEZ, D. F., LOPEZ-ROMERO, E., ZINKER, S., PONCE-NOYOLA, P., KULLBERG, B. J., BROWN, A. J., ODDS, F. C., FLORES-CARREON, A. & GOW, N. A. 2007. Endoplasmic reticulum alpha-glycosidases of *Candida albicans* are required for N glycosylation, cell wall integrity, and normal host-fungus interaction. *Eukaryot Cell*, 6, 2184-93.
- MOYES, D. L. & NAGLIK, J. R. 2011. Mucosal immunity and *Candida albicans* infection. *Clinical & developmental immunology*, 2011, 346307-346307.
- MOYES, D. L., WILSON, D., RICHARDSON, J. P., MOGAVERO, S., TANG, S. X., WERNECKE, J., HOF, S., GRATACAP, R. L., ROBBINS, J., RUNGLALL, M., MURCIANO, C., BLAGOJEVIC, M., THAVARAJ, S., FORSTER, T. M., HEBECKER, B., KASPER, L., VIZCAY, G., IANCU, S. I., KICHIK, N., HADER, A., KURZAI, O., LUO, T., KRUGER, T., KNIEMEYER, O., COTA, E., BADER, O., WHEELER, R. T., GUTSMANN, T., HUBE, B. & NAGLIK, J. R. 2016. Candidalysin is a fungal peptide toxin critical for mucosal infection. *Nature*, 532, 64-8.
- MUADCHEINGKA, T. & TANTIVITAYAKUL, P. 2015. Distribution of *Candida albicans* and non-*albicans* *Candida* species in oral candidiasis patients: Correlation between cell surface hydrophobicity and biofilm forming activities. *Arch Oral Biol*, 60, 894-901.
- MUKHERJEE, P. K., WANG, H., RETUERTO, M., ZHANG, H., BURKEY, B., GHANNOUM, M. A. & ENG, C. 2017. Bacteriome and mycobiome associations in oral tongue cancer. *Oncotarget*, 8, 97273-97289.
- NAGLIK, J. R., MOYES, D., MAKWANA, J., KANZARIA, P., TSICHLAKI, E., WEINDL, G., TAPPUNI, A. R., RODGERS, C. A., WOODMAN, A. J., CHALLACOMBE, S. J., SCHALLER, M. & HUBE, B. 2008. Quantitative expression of the *Candida albicans* secreted aspartyl proteinase gene family in human oral and vaginal candidiasis. *Microbiology*, 154, 3266-3280.
- NAGLIK, J. R., RODGERS, C. A., SHIRLAW, P. J., DOBBIE, J. L., FERNANDES-NAGLIK, L. L., GREENSPAN, D., AGABIAN, N. & CHALLACOMBE, S. J. 2003. Differential expression of *Candida albicans* secreted aspartyl proteinase and phospholipase B genes in humans correlates with active oral and vaginal infections. *J Infect Dis*, 188, 469-79.
- NAIR, R. G., ANIL, S. & SAMARANAYAKE, L. P. 2001. The effect of oral bacteria on *Candida albicans* germ-tube formation. *APMIS*, 109, 147-54.
- NAKAMURA, Y., YAMAMOTO, N., KINO, Y., YAMAMOTO, N., KAMEI, S., MORI, H., KUROKAWA, K. & NAKASHIMA, N. 2016. Establishment of a multi-species biofilm

References

- model and metatranscriptomic analysis of biofilm and planktonic cell communities. *Appl Microbiol Biotechnol*, 100, 7263-79.
- NETT, J., LINCOLN, L., MARCHILLO, K., MASSEY, R., HOLOYDA, K., HOFF, B., VANHANDEL, M. & ANDES, D. 2007. Putative role of beta-1,3 glucans in *Candida albicans* biofilm resistance. *Antimicrob Agents Chemother*, 51, 510-20.
- NETT, J. E. 2016. The Host's Reply to *Candida* Biofilm. *Pathogens*, 5.
- NETT, J. E., CRAWFORD, K., MARCHILLO, K. & ANDES, D. R. 2010a. Role of Fks1p and matrix glucan in *Candida albicans* biofilm resistance to an echinocandin, pyrimidine, and polyene. *Antimicrob Agents Chemother*, 54, 3505-8.
- NETT, J. E., SANCHEZ, H., CAIN, M. T. & ANDES, D. R. 2010b. Genetic basis of *Candida* biofilm resistance due to drug-sequestering matrix glucan. *J Infect Dis*, 202, 171-5.
- NG, T. S., DESA, M. N. M., SANDAI, D., CHONG, P. P. & THAN, L. T. L. 2016. Growth, biofilm formation, antifungal susceptibility and oxidative stress resistance of *Candida glabrata* are affected by different glucose concentrations. *Infect Genet Evol*, 40, 331-338.
- NOBILE, C. 2013. The Role of *Candida albicans* Biofilms in Human Disease.
- NOBILE, C. J., FOX, E. P., NETT, J. E., SORRELLS, T. R., MITROVICH, Q. M., HERNDAY, A. D., TUCH, B. B., ANDES, D. R. & JOHNSON, A. D. 2012. A recently evolved transcriptional network controls biofilm development in *Candida albicans*. *Cell*, 148, 126-38.
- NOBILE, C. J. & JOHNSON, A. D. 2015. *Candida albicans* Biofilms and Human Disease. *Annu Rev Microbiol*, 69, 71-92.
- NOBILE, C. J. & MITCHELL, A. P. 2005. Regulation of cell-surface genes and biofilm formation by the *C. albicans* transcription factor Bcr1p. *Curr Biol*, 15, 1150-5.
- NOVERR, M. C. & HUFFNAGLE, G. B. 2004. Regulation of *Candida albicans* morphogenesis by fatty acid metabolites. *Infect Immun*, 72, 6206-10.
- NUNES, E. M., POLICASTRO, V. B., SCAVASSIN, P. M., LEITE, A. R., MENDOZA MARIN, D. O., GIRO, G., DE OLIVEIRA JUNIOR, N. M., COMPAGNONI, M. A. & PERO, A. C. 2016. Crossover clinical trial of different methods of removing a denture adhesive and the influence on the oral microbiota. *J Prosthet Dent*, 115, 462-8.
- NYGAARD, A. B., TUNSIJØ, H. S., MEISAL, R. & CHARNOCK, C. 2020. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. *Scientific Reports*, 10, 3209.
- O'CONNELL, L. M., SANTOS, R., SPRINGER, G., BURNE, R. A., NASCIMENTO, M. M. & RICHARDS, V. P. 2020. Site-Specific Profiling of the Dental Mycobiome Reveals Strong Taxonomic Shifts during Progression of Early-Childhood Caries. *Appl Environ Microbiol*, 86.
- O'DONNELL, L. E., ALALWAN, H. K., KEAN, R., CALVERT, G., NILE, C. J., LAPPIN, D. F., ROBERTSON, D., WILLIAMS, C., RAMAGE, G. & SHERRY, L. 2017. *Candida albicans* biofilm heterogeneity does not influence denture stomatitis but strongly influences denture cleansing capacity. *J Med Microbiol*, 66, 54-60.
- O'DONNELL, L. E., MILLHOUSE, E., SHERRY, L., KEAN, R., MALCOLM, J., NILE, C. J. & RAMAGE, G. 2015a. Polymicrobial *Candida* biofilms: friends and foe in the oral cavity. *FEMS Yeast Res*, 15.
- O'DONNELL, L. E., ROBERTSON, D., NILE, C. J., CROSS, L. J., RIGGIO, M., SHERRIFF, A., BRADSHAW, D., LAMBERT, M., MALCOLM, J., BUIJS, M. J., ZAURA, E., CRIELAARD, W., BRANDT, B. W. & RAMAGE, G. 2015b. The Oral Microbiome of Denture Wearers Is Influenced by Levels of Natural Dentition. *PLoS One*, 10, e0137717.

References

- O'DONNELL, L. E., SMITH, K., WILLIAMS, C., NILE, C. J., LAPPIN, D. F., BRADSHAW, D., LAMBERT, M., ROBERTSON, D. P., BAGG, J., HANNAH, V. & RAMAGE, G. 2015c. Dentures are a Reservoir for Respiratory Pathogens. *J Prosthodont*.
- O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBBERTSE, B., SMITH-WHITE, B., AKO-ADJEI, D., ASTASHYN, A., BADRETDIN, A., BAO, Y., BLINKOVA, O., BROVER, V., CHETVERNIN, V., CHOI, J., COX, E., ERMOLAEVA, O., FARRELL, C. M., GOLDFARB, T., GUPTA, T., HAFT, D., HATCHER, E., HLAVINA, W., JOARDAR, V. S., KODALI, V. K., LI, W., MAGLOTT, D., MASTERSON, P., MCGARVEY, K. M., MURPHY, M. R., O'NEILL, K., PUJAR, S., RANGWALA, S. H., RAUSCH, D., RIDDICK, L. D., SCHOCH, C., SHKEDA, A., STORZ, S. S., SUN, H., THIBAUD-NISSEN, F., TOLSTOY, I., TULLY, R. E., VATSAN, A. R., WALLIN, C., WEBB, D., WU, W., LANDRUM, M. J., KIMCHI, A., TATUSOVA, T., DICUCCIO, M., KITTS, P., MURPHY, T. D. & PRUITT, K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44, D733-45.
- O'DONNELL, L. E., ALALWAN, H. K. A., KEAN, R., CALVERT, G., NILE, C. J., LAPPIN, D. F., ROBERTSON, D., WILLIAMS, C., RAMAGE, G. & SHERRY, L. 2017. Candida albicans biofilm heterogeneity does not influence denture stomatitis but strongly influences denture cleansing capacity. *Journal of Medical Microbiology*, 66, 54-60.
- O'DONNELL, L. E., ROBERTSON, D., NILE, C. J., CROSS, L. J., RIGGIO, M., SHERRIFF, A., BRADSHAW, D., LAMBERT, M., MALCOLM, J., BUIJS, M. J., ZAURA, E., CRIELAARD, W., BRANDT, B. W. & RAMAGE, G. 2015. The Oral Microbiome of Denture Wearers Is Influenced by Levels of Natural Dentition. *PLOS ONE*, 10, e0137717.
- OKONECHNIKOV, K., CONESA, A. & GARCÍA-ALCALDE, F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 32, 292-294.
- OZCAN, K., ILKIT, M., ATEŞ, A., TURAC-BICER, A. & DEMIRHINDI, H. 2010. Performance of Chromogenic Candida Agar and CHROMagar Candida in recovery and presumptive identification of monofungal and polyfungal vaginal isolates. *Medical Mycology*, 48, 29-34.
- PAGE, R. C., ENGEL, L. D., NARAYANAN, A. S. & CLAGETT, J. A. 1978. Chronic inflammatory gingival and periodontal disease. *JAMA*, 240, 545-50.
- PANTAZOPOULOU, A. & DIALLINAS, G. 2007. Fungal nucleobase transporters. *FEMS Microbiology Reviews*, 31, 657-675.
- PARKS, D. H., RINKE, C., CHUVOCHINA, M., CHAUMEIL, P.-A., WOODCROFT, B. J., EVANS, P. N., HUGENHOLTZ, P. & TYSON, G. W. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2, 1533-1542.
- PAROLIN, C., MARANGONI, A., LAGHI, L., FOSCHI, C., NAHUI PALOMINO, R. A., CALONGHI, N., CEVENINI, R. & VITALI, B. 2015. Isolation of Vaginal Lactobacilli and Characterization of Anti-Candida Activity. *PLoS One*, 10, e0131220.
- PARTOAZAR, A., TALAEI, N., BAHADOR, A., POURHAJIBAGHER, M., DEHPUR, S., SADATI, M. & BAKHTIARIAN, A. 2019. Antibiofilm activity of natural zeolite supported NanoZnO: inhibition of Esp gene expression of Enterococcus faecalis. *Nanomedicine (Lond)*, 14, 675-687.
- PASTER, B. J., OLSEN, I., AAS, J. A. & DEWHIRST, F. E. 2006. The breadth of bacterial diversity in the human periodontal pocket and other oral sites. *Periodontol 2000*, 42, 80-7.

References

- PAWLAK, D., SCHIELMANN, M., WOJCIECHOWSKI, M. & ANDRUSZKIEWICZ, R. 2016. Synthesis and biological activity of novel ester derivatives of N(3)-(4-metoxymethyl)-(S)-2,3-diaminopropanoic acid containing amide and keto function as inhibitors of glucosamine-6-phosphate synthase. *Bioorg Med Chem Lett*, 26, 3586-9.
- PELEG, A. Y., HOGAN, D. A. & MYLONAKIS, E. 2010. Medically important bacterial–fungal interactions. *Nature Reviews Microbiology*, 8, 340-349.
- PEMMARAJU, S. C., PRUTHI, P. A., PRASAD, R. & PRUTHI, V. 2016. Modulation of *Candida albicans* Biofilm by Different Carbon Sources. *Mycopathologia*, 181, 341-52.
- PERLROTH, J., CHOI, B. & SPELLBERG, B. 2007. Nosocomial fungal infections: epidemiology, diagnosis, and treatment. *Med Mycol*, 45, 321-46.
- PERSOON, I. F., BUIJS, M. J., OZOK, A. R., CRIELAARD, W., KROM, B. P., ZAURA, E. & BRANDT, B. W. 2017a. The mycobiome of root canal infections is correlated to the bacteriome. *Clin Oral Investig*, 21, 1871-1881.
- PERSOON, I. F., BUIJS, M. J., ÖZOK, A. R., CRIELAARD, W., KROM, B. P., ZAURA, E. & BRANDT, B. W. 2017b. The mycobiome of root canal infections is correlated to the bacteriome. *Clinical Oral Investigations*, 21, 1871-1881.
- PETERS, B. A., WU, J., HAYES, R. B. & AHN, J. 2017. The oral fungal mycobiome: characteristics and relation to periodontitis in a pilot study. *BMC Microbiol*, 17, 157.
- PETERS, B. M., JABRA-RIZK, M. A., O'MAY, G. A., COSTERTON, J. W. & SHIRTLIFF, M. E. 2012. Polymicrobial interactions: impact on pathogenesis and human disease. *Clin Microbiol Rev*, 25, 193-213.
- PETERS, B. M. & NOVERR, M. C. 2013. *Candida albicans*-*Staphylococcus aureus* polymicrobial peritonitis modulates host innate immunity. *Infect Immun*, 81, 2178-89.
- PIERCE, C. G., CHATURVEDI, A. K., LAZZELL, A. L., POWELL, A. T., SAVILLE, S. P., MCHARDY, S. F. & LOPEZ-RIBOT, J. L. 2015. A Novel Small Molecule Inhibitor of *Candida albicans* Biofilm Formation, Filamentation and Virulence with Low Potential for the Development of Resistance. *NPJ Biofilms Microbiomes*, 1, 15012-.
- PITT, J. J. 2009. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev*, 30, 19-34.
- PONGRACZ, J., BENEDEK, K., JUHASZ, E., IVAN, M. & KRISTOF, K. 2016. In vitro biofilm production of *Candida* bloodstream isolates: any association with clinical characteristics? *J Med Microbiol*, 65, 272-7.
- POTERA, C. 1999. Forging a link between biofilms and disease. *Science*, 283, 1837, 1839.
- POULAIN, D. 2015. *Candida albicans*, plasticity and pathogenesis. *Crit Rev Microbiol*, 41, 208-17.
- PROCTOR, L., LOTEMPIO, J., MARQUITZ, A., DASCHNER, P., XI, D., FLORES, R., BROWN, L., RANALLO, R., MARUVADA, P., REGAN, K., DWAYNE LUNSFORD, R., REDDY, M., CALER, L. & TEAM, N. I. H. H. M. P. A. 2019. A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. *Microbiome*, 7, 31.
- PRUESSE, E., QUAST, C., KNITTEL, K., FUCHS, B. M., LUDWIG, W., PEPLIES, J. & GLOCKNER, F. O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*, 35, 7188-96.
- RAJENDRAN, R., MAY, A., SHERRY, L., KEAN, R., WILLIAMS, C., JONES, B. L., BURGESS, K. V., HERINGA, J., ABELN, S., BRANDT, B. W., MUNRO, C. A. & RAMAGE, G. 2016a.

References

- Integrating *Candida albicans* metabolism with biofilm heterogeneity by transcriptome mapping. *Scientific Reports*, 6, 35436.
- RAJENDRAN, R., MAY, A., SHERRY, L., KEAN, R., WILLIAMS, C., JONES, B. L., BURGESS, K. V., HERINGA, J., ABELN, S., BRANDT, B. W., MUNRO, C. A. & RAMAGE, G. 2016b. Integrating *Candida albicans* metabolism with biofilm heterogeneity by transcriptome mapping. *Sci Rep*, 6, 35436.
- RAJENDRAN, R., SHERRY, L., DESHPANDE, A., JOHNSON, E. M., HANSON, M. F., WILLIAMS, C., MUNRO, C. A., JONES, B. L. & RAMAGE, G. 2016c. A Prospective Surveillance Study of Candidaemia: Epidemiology, Risk Factors, Antifungal Treatment and Outcome in Hospitalized Patients. *Front Microbiol*, 7, 915.
- RAJENDRAN, R., SHERRY, L., LAPPIN, D. F., NILE, C. J., SMITH, K., WILLIAMS, C., MUNRO, C. A. & RAMAGE, G. 2014. Extracellular DNA release confers heterogeneity in *Candida albicans* biofilm formation. *BMC Microbiol*, 14, 303.
- RAJENDRAN, R., SHERRY, L., NILE, C. J., SHERRIFF, A., JOHNSON, E., HANSON, M., WILLIAMS, C., MUNRO, C., JONES, B. & RAMAGE, G. 2015. Biofilm formation is a risk factor for mortality in patients with *Candida albicans* bloodstream infection - Scotland, 2012-2013. *Clin Microbiol Infect*.
- RAJENDRAN, R., SHERRY, L., NILE, C. J., SHERRIFF, A., JOHNSON, E. M., HANSON, M. F., WILLIAMS, C., MUNRO, C. A., JONES, B. J. & RAMAGE, G. 2016d. Biofilm formation is a risk factor for mortality in patients with *Candida albicans* bloodstream infection-Scotland, 2012-2013. *Clin Microbiol Infect*, 22, 87-93.
- RAMACHANDRA, S., LINDE, J., BROCK, M., GUTHKE, R., HUBE, B. & BRUNKE, S. 2014. Regulatory networks controlling nitrogen sensing and uptake in *Candida albicans*. *PLoS One*, 9, e92734.
- RAMAGE, G., COCO, B., SHERRY, L., BAGG, J. & LAPPIN, D. F. 2012a. In vitro *Candida albicans* biofilm induced proteinase activity and SAP8 expression correlates with in vivo denture stomatitis severity. *Mycopathologia*, 174, 11-19.
- RAMAGE, G., LAPPIN, D. F., MILLHOUSE, E., MALCOLM, J., JOSE, A., YANG, J., BRADSHAW, D. J., PRATTEN, J. R. & CULSHAW, S. 2017. The epithelial cell response to health and disease associated oral biofilm models. *J Periodontal Res*, 52, 325-333.
- RAMAGE, G., O'DONNELL, L., SHERRY, L., CULSHAW, S., BAGG, J., CZESNIKIEWICZ-GUZIŁ, M., BROWN, C., MCKENZIE, D., CROSS, L., MACINNES, A., BRADSHAW, D., VARGHESE, R., GOMEZ PEREIRA, P., JOSE, A., SANYAL, S. & ROBERTSON, D. 2019. Impact of frequency of denture cleaning on microbial and clinical parameters - a bench to chairside approach. *J Oral Microbiol*, 11, 1538437.
- RAMAGE, G., RAJENDRAN, R., SHERRY, L. & WILLIAMS, C. 2012b. Fungal biofilm resistance. *Int J Microbiol*, 2012, 528521.
- RAMAGE, G., ROBERTSON, S. N. & WILLIAMS, C. 2014. Strength in numbers: antifungal strategies against fungal biofilms. *Int J Antimicrob Agents*, 43, 114-20.
- RAMAGE, G., VANDE WALLE, K., WICKES, B. L. & LOPEZ-RIBOT, J. L. 2001a. Biofilm formation by *Candida dubliniensis*. *J Clin Microbiol*, 39, 3234-40.
- RAMAGE, G., VANDE WALLE, K., WICKES, B. L. & LOPEZ-RIBOT, J. L. 2001b. Standardized method for in vitro antifungal susceptibility testing of *Candida albicans* biofilms. *Antimicrob Agents Chemother*, 45, 2475-9.
- RAMAGE, G., VANDEWALLE, K., LOPEZ-RIBOT, J. L. & WICKES, B. L. 2002a. The filamentation pathway controlled by the Efg1 regulator protein is required for normal biofilm formation and development in *Candida albicans*. *FEMS Microbiol Lett*, 214, 95-100.

References

- RAMAGE, G., VANDEWALLE, K., LÓPEZ-RIBOT, J. L. & WICKES, B. L. 2002b. The filamentation pathway controlled by the Efg1 regulator protein is required for normal biofilm formation and development in *Candida albicans*. *FEMS Microbiology Letters*, 214, 95-100.
- RAMAGE, G. & WILLIAMS, C. 2012. Fungal biofilms.
- RAMAGE, G., ZALEWSKA, A., CAMERON, D. A., SHERRY, L., MURRAY, C., FINNEGAN, M. B., LOEWY, Z. G. & JAGGER, D. C. 2012c. A comparative in vitro study of two denture cleaning techniques as an effective strategy for inhibiting *Candida albicans* biofilms on denture surfaces and reducing inflammation. *J Prosthodont*, 21, 516-22.
- RICHARDSON, M., BOWYER, P. & SABINO, R. 2019. The human lung and *Aspergillus*: You are what you breathe in? *Medical Mycology*, 57, S145-S154.
- ROCAS, I. N., SIQUEIRA, J. F., JR., SANTOS, K. R. & COELHO, A. M. 2001. "Red complex" (*Bacteroides forsythus*, *Porphyromonas gingivalis*, and *Treponema denticola*) in endodontic infections: a molecular approach. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, 91, 468-71.
- ROHART, F., GAUTIER, B., SINGH, A. & KA, L. C. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*, 13, e1005752.
- ROILIDES, E. 2016. Emerging fungi causing human infection: new or better identified? *Clin Microbiol Infect*, 22, 660-1.
- ROMERO-LASTRA, P., SÁNCHEZ, M. C., LLAMA-PALACIOS, A., FIGUERO, E., HERRERA, D. & SANZ, M. 2019. Gene expression of *Porphyromonas gingivalis* ATCC 33277 when growing in an in vitro multispecies biofilm. *PLOS ONE*, 14, e0221234.
- ROMO, J. A., ZHANG, H., CAI, H., KADOSH, D., KOEHLER, J. R., SAVILLE, S. P., WANG, Y. & LOPEZ-RIBOT, J. L. 2019. Global Transcriptomic Analysis of the *Candida albicans* Response to Treatment with a Novel Inhibitor of Filamentation. *mSphere*, 4.
- ROSSELL, D., STEPHAN-OTTO ATTOLINI, C., KROISS, M. & STÖCKER, A. 2014. QUANTIFYING ALTERNATIVE SPLICING FROM PAIRED-END RNA-SEQUENCING DATA. *Ann Appl Stat*, 8, 309-330.
- SACHDEO, A., HAFFAJEE, A. D. & SOCRANSKY, S. S. 2008. Biofilms in the edentulous oral cavity. *J Prosthodont*, 17, 348-56.
- SADIG, W. 2010. The denture hygiene, denture stomatitis and role of dental hygienist. *Int J Dent Hyg*, 8, 227-31.
- SANTANA, I. L., GONÇALVES, L. M., DE VASCONCELLOS, A. A., DA SILVA, W. J., CURY, J. A. & DEL BEL CURY, A. A. 2013. Dietary carbohydrates modulate *Candida albicans* biofilm development on the denture surface. *PLoS One*, 8, e64645.
- SAUER, U. 2006. Metabolic networks in motion: 13C-based flux analysis. *Mol Syst Biol*, 2, 62.
- SCHAUER, N., STEINHAUSER, D., STRELKOV, S., SCHOMBURG, D., ALLISON, G., MORITZ, T., LUNDGREN, K., ROESSNER-TUNALI, U., FORBES, M. G., WILLMITZER, L., FERNIE, A. R. & KOPKA, J. 2005. GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett*, 579, 1332-7.
- SETTEM, R. P., EL-HASSAN, A. T., HONMA, K., STAFFORD, G. P. & SHARMA, A. 2012. *Fusobacterium nucleatum* and *Tannerella forsythia* induce synergistic alveolar bone loss in a mouse periodontitis model. *Infect Immun*, 80, 2436-43.
- SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13, 2498-2504.

References

- SHERRY, L., KEAN, R., MCKLOUD, E., O'DONNELL, L. E., METCALFE, R., JONES, B. L. & RAMAGE, G. 2017. Biofilms Formed by Isolates from Recurrent Vulvovaginal Candidiasis Patients Are Heterogeneous and Insensitive to Fluconazole. *Antimicrob Agents Chemother*, 61.
- SHERRY, L., LAPPIN, G., O'DONNELL, L. E., MILLHOUSE, E., MILLINGTON, O. R., BRADSHAW, D. J., AXE, A. S., WILLIAMS, C., NILE, C. J. & RAMAGE, G. 2016a. Viable Compositional Analysis of an Eleven Species Oral Polymicrobial Biofilm. *Front Microbiol*, 7, 912.
- SHERRY, L., LAPPIN, G., O'DONNELL, L. E., MILLHOUSE, E., MILLINGTON, O. R., BRADSHAW, D. J., AXE, A. S., WILLIAMS, C., NILE, C. J. & RAMAGE, G. 2016b. Viable Compositional Analysis of an Eleven Species Oral Polymicrobial Biofilm. *Frontiers in Microbiology*, 7.
- SHERRY, L., RAJENDRAN, R., LAPPIN, D. F., BORGHİ, E., PERDONI, F., FALLENI, M., TOSI, D., SMITH, K., WILLIAMS, C., JONES, B., NILE, C. J. & RAMAGE, G. 2014. Biofilms formed by *Candida albicans* bloodstream isolates display phenotypic and transcriptional heterogeneity that are associated with resistance and pathogenicity. *BMC Microbiol*, 14, 182.
- SHI, B., WU, T., MCLEAN, J., EDLUND, A., YOUNG, Y., HE, X., LV, H., ZHOU, X., SHI, W., LI, H. & LUX, R. 2016. The Denture-Associated Oral Microbiome in Health and Stomatitis. *mSphere*, 1.
- SHIBASAKI, S., KARASAKI, M., AOKI, W. & UEDA, M. 2018. Molecular and Physiological Study of *Candida albicans* by Quantitative Proteome Analysis. *Proteomes*, 6, 34.
- SHIBLI, J. A., MELO, L., FERRARI, D. S., FIGUEIREDO, L. C., FAVERI, M. & FERES, M. 2008. Composition of supra- and subgingival biofilm of subjects with healthy and diseased implants. *Clin Oral Implants Res*, 19, 975-82.
- SILAO, F. G. S., WARD, M., RYMAN, K., WALLSTRÖM, A., BRINDEFALK, B., UDEKWU, K. & LJUNGDAHL, P. O. 2019. Mitochondrial proline catabolism activates Ras1/cAMP/PKA-induced filamentation in *Candida albicans*. *PLoS Genet*, 15, e1007976.
- SILVA, S., HENRIQUES, M., HAYES, A., OLIVEIRA, R., AZEREDO, J. & WILLIAMS, D. W. 2011. *Candida glabrata* and *Candida albicans* co-infection of an in vitro oral epithelium. *Journal of Oral Pathology & Medicine*, 40, 421-427.
- SINTIM, H. O. & GURSOY, U. K. 2016. Biofilms as "Connectors" for Oral and Systems Medicine: A New Opportunity for Biomarkers, Molecular Targets, and Bacterial Eradication. *OMICS*, 20, 3-11.
- SKRZYPEK, M. S., BINKLEY, J., BINKLEY, G., MIYASATO, S. R., SIMISON, M. & SHERLOCK, G. 2016. The *Candida* Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Research*, 45, D592-D596.
- SLENTER, D. N., KUTMON, M., HANSPERS, K., RIUTTA, A., WINDSOR, J., NUNES, N., MÉLIUS, J., CIRILLO, E., COORT, S. L., DIGLES, D., EHRHART, F., GIESBERTZ, P., KALAFATI, M., MARTENS, M., MILLER, R., NISHIDA, K., RIESWIJK, L., WAAGMEESTER, A., EIJSSEN, L. M. T., EVELO, C. T., PICO, A. R. & WILLIGHAGEN, E. L. 2018. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res*, 46, D661-d667.
- SOBEL, J. B., GOLDFARB, I. W., SLATER, H. & HAMMELL, E. J. 1992. Inhalation injury: a decade without progress. *J Burn Care Rehabil*, 13, 573-5.
- SOCRANSKY, S. S., HAFFAJEE, A. D., CUGINI, M. A., SMITH, C. & KENT, R. L., JR. 1998. Microbial complexes in subgingival plaque. *J Clin Periodontol*, 25, 134-44.

References

- SOKOL, H., LEDUCQ, V., ASCHARD, H., PHAM, H. P., JEGOU, S., LANDMAN, C., COHEN, D., LIGUORI, G., BOURRIER, A., NION-LARMURIER, I., COSNES, J., SEKSIK, P., LANGELLA, P., SKURNIK, D., RICHARD, M. L. & BEAUGERIE, L. 2017. Fungal microbiota dysbiosis in IBD. *Gut*, 66, 1039-1048.
- SOLDINI, S., POSTERARO, B., VELLA, A., DE CAROLIS, E., BORGHI, E., FALLENI, M., LOSITO, A. R., MAIURO, G., TRECARICHI, E. M., SANGUINETTI, M. & TUMBARELLO, M. 2017. Microbiological and clinical characteristics of biofilm-forming *Candida parapsilosis* isolates associated with fungaemia and their impact on mortality. *Clin Microbiol Infect*.
- SOUSA, V., NIBALI, L., SPRATT, D., DOPICO, J., MARDAS, N., PETRIE, A. & DONOS, N. 2017. Peri-implant and periodontal microbiome diversity in aggressive periodontitis patients: a pilot study. *Clin Oral Implants Res*, 28, 558-570.
- STEIN-O'BRIEN, G. L., ARORA, R., CULHANE, A. C., FAVOROV, A. V., GARMIRE, L. X., GREENE, C. S., GOFF, L. A., LI, Y., NGOM, A., OCHS, M. F., XU, Y. & FERTIG, E. J. 2018. Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet*, 34, 790-805.
- STIPETIC, L. H., DALBY, M. J., DAVIES, R. L., MORTON, F. R., RAMAGE, G. & BURGESS, K. E. V. 2016. A novel metabolomic approach used for the comparison of *Staphylococcus aureus* planktonic cells and biofilm samples. *Metabolomics*, 12, 75.
- STRIJBIS, K. & DISTEL, B. 2010. Intracellular acetyl unit transport in fungal carbon metabolism. *Eukaryot Cell*, 9, 1809-15.
- STRUZYCKA, I. 2014. The oral microbiome in dental caries. *Pol J Microbiol*, 63, 127-35.
- SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102, 15545-50.
- SUDBERY, P. E. 2011. Growth of *Candida albicans* hyphae. *Nat Rev Microbiol*, 9, 737-48.
- SUE, T., OBOLONKIN, V., GRIFFITHS, H. & VILLAS-BÔAS, S. G. 2011. An Exometabolomics Approach to Monitoring Microbial Contamination in Microalgal Fermentation Processes by Using Metabolic Footprint Analysis. *Applied and Environmental Microbiology*, 77, 7605.
- SZKLARCZYK, D., GABLE, A. L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N. T., MORRIS, J. H., BORK, P., JENSEN, L. J. & MERING, C. V. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47, D607-d613.
- SZTUKOWSKA, M. N., DUTTON, L. C., DELANEY, C., RAMSDALE, M., RAMAGE, G., JENKINSON, H. F., NOBBS, A. H. & LAMONT, R. J. 2018a. Community Development between *Porphyromonas gingivalis* and *Candida albicans* Mediated by InlJ and Als3. *mBio*.
- SZTUKOWSKA, M. N., DUTTON, L. C., DELANEY, C., RAMSDALE, M., RAMAGE, G., JENKINSON, H. F., NOBBS, A. H. & LAMONT, R. J. 2018b. Community Development between *Porphyromonas gingivalis* and *Candida albicans* Mediated by InlJ and Als3. *mBio*, 9.
- TAFF, H. T., NETT, J. E. & ANDES, D. R. 2012. Comparative analysis of *Candida* biofilm quantitation assays. *Med Mycol*, 50, 214-8.
- TAMAI, R., SUGAMATA, M. & KIYOURA, Y. 2011. *Candida albicans* enhances invasion of human gingival epithelial cells and gingival fibroblasts by *Porphyromonas gingivalis*. *Microb Pathog*, 51, 250-4.

References

- TAMMER, I., REUNER, J., HARTIG, R. & GEGINAT, G. 2014. Induction of *Candida albicans* biofilm formation on silver-coated vascular grafts. *J Antimicrob Chemother*, 69, 1282-5.
- TAN, X., QIN, N., WU, C., SHENG, J., YANG, R., ZHENG, B., MA, Z., LIU, L., PENG, X. & JIA, A. 2015. Transcriptome analysis of the biofilm formed by methicillin-susceptible *Staphylococcus aureus*. *Sci Rep*, 5, 11997.
- TANG, J. 2011. Microbial metabolomics. *Curr Genomics*, 12, 391-403.
- TAO, L., ZHANG, Y., FAN, S., NOBILE, C. J., GUAN, G. & HUANG, G. 2017. Integration of the tricarboxylic acid (TCA) cycle with cAMP signaling and Sfl2 pathways in the regulation of CO₂ sensing and hyphal development in *Candida albicans*. *PLoS Genet*, 13, e1006949.
- TATUSOVA, T., CIUFO, S., FEDERHEN, S., FEDOROV, B., MCVEIGH, R., O'NEILL, K., TOLSTOY, I. & ZASLAVSKY, L. 2014. Update on RefSeq microbial genomes resources. *Nucleic acids research*, 43.
- TELES, F. R., TELES, R. P., SACHDEO, A., UZEL, N. G., SONG, X. Q., TORRESYAP, G., SINGH, M., PAPAS, A., HAFFAJEE, A. D. & SOCRANSKY, S. S. 2012. Comparison of microbial changes in early redeveloping biofilms on natural teeth and dentures. *J Periodontol*, 83, 1139-48.
- THEILADE, E. 1986. The non-specific theory in microbial etiology of inflammatory periodontal diseases. *J Clin Periodontol*, 13, 905-11.
- THEIN, Z. M., SAMARANAYAKE, Y. H. & SAMARANAYAKE, L. P. 2006. Effect of oral bacteria on growth and survival of *Candida albicans* biofilms. *Arch Oral Biol*, 51, 672-80.
- TIEW, P. Y., MAC AOGAIN, M., ALI, N. A. T. B. M., THNG, K. X., GOH, K., LAU, K. J. X. & CHOTIRMALL, S. H. 2020. The Mycobiome in Health and Disease: Emerging Concepts, Methodologies and Challenges. *Mycopathologia*, 185, 207-231.
- TRIPATHI, G., WILTSHIRE, C., MACASKILL, S., TOURNU, H., BUDGE, S. & BROWN, A. J. 2002. Gcn4 co-ordinates morphogenetic and metabolic responses to amino acid starvation in *Candida albicans*. *EMBO J*, 21, 5448-56.
- TUMBARELLO, M., FIORI, B., TRECARCHI, E. M., POSTERARO, P., LOSITO, A. R., DE LUCA, A., SANGUINETTI, M., FADDA, G., CAUDA, R. & POSTERARO, B. 2012. Risk factors and outcomes of candidemia caused by biofilm-forming isolates in a tertiary care hospital. *PLoS One*, 7, e33705.
- TUMBARELLO, M., POSTERARO, B., TRECARCHI, E. M., FIORI, B., ROSSI, M., PORTA, R., DE GAETANO DONATI, K., LA SORDA, M., SPANU, T., FADDA, G., CAUDA, R. & SANGUINETTI, M. 2007. Biofilm production by *Candida* species and inadequate antifungal therapy as predictors of mortality for patients with candidemia. *J Clin Microbiol*, 45, 1843-50.
- UPPULURI, P., CHATURVEDI, A. K., SRINIVASAN, A., BANERJEE, M., RAMASUBRAMANIAM, A. K., KOHLER, J. R., KADOSH, D. & LOPEZ-RIBOT, J. L. 2010. Dispersion as an important step in the *Candida albicans* biofilm developmental cycle. *PLoS Pathog*, 6, e1000828.
- UWAMAHORO, N., VERMA-GAUR, J., SHEN, H. H., QU, Y., LEWIS, R., LU, J., BAMBERY, K., MASTERS, S. L., VINCE, J. E., NADERER, T. & TRAVEN, A. 2014. The pathogen *Candida albicans* hijacks pyroptosis for escape from macrophages. *mBio*, 5, e00003-14.
- VIEIRA COLOMBO, A. P., MAGALHAES, C. B., HARTENBACH, F. A., MARTINS DO SOUTO, R. & MACIEL DA SILVA-BOGHOSSIAN, C. 2016. Periodontal-disease-associated biofilm: A reservoir for pathogens of medical importance. *Microb Pathog*, 94, 27-34.

References

- VILLA, S., HAMIDEH, M., WEINSTOCK, A., QASIM, M. N., HAZBUN, T. R., SELLAM, A., HERNDAY, A. D. & THANGAMANI, S. 2020. Transcriptional control of hyphal morphogenesis in *Candida albicans*. *FEMS Yeast Res*, 20.
- VILLAS-BOAS, S. G., MAS, S., AKESSON, M., SMEDSGAARD, J. & NIELSEN, J. 2005. Mass spectrometry in metabolome analysis. *Mass Spectrom Rev*, 24, 613-46.
- VYLKOVA, S., CARMAN, A. J., DANHOF, H. A., COLLETTE, J. R., ZHOU, H. & LORENZ, M. C. 2011a. The Fungal Pathogen *Candida albicans* Autoinduces Hyphal Morphogenesis by Raising Extracellular pH. *mBio*, 2, e00055-11.
- VYLKOVA, S., CARMAN, A. J., DANHOF, H. A., COLLETTE, J. R., ZHOU, H. & LORENZ, M. C. 2011b. The fungal pathogen *Candida albicans* autoinduces hyphal morphogenesis by raising extracellular pH. *mBio*, 2, e00055-11.
- WALKER, L. A. & MUNRO, C. A. 2020. Caspofungin Induced Cell Wall Changes of *Candida* Species Influences Macrophage Interactions. *Frontiers in Cellular and Infection Microbiology*, 10.
- WAN, N., WANG, H., NG, C. K., MUKHERJEE, M., REN, D., CAO, B. & TANG, Y. J. 2018. Bacterial Metabolism During Biofilm Growth Investigated by ¹³C Tracing. *Frontiers in Microbiology*, 9.
- WASHIO, J., MAYANAGI, G. & TAKAHASHI, N. 2010. Challenge to Metabolomics of Oral Biofilm: —From “What Are They?” to “What Are They Doing?”—. *Journal of Oral Biosciences*, 52, 225-232.
- WECKWERTH, W. 2010. Metabolomics: an integral technique in systems biology. *Bioanalysis*, 2, 829-36.
- WEERASEKERA, M. M., WIJESINGHE, G. K., JAYARATHNA, T. A., GUNASEKARA, C. P., FERNANDO, N., KOTTEGODA, N. & SAMARANAYAKE, L. P. 2016. Culture media profoundly affect *Candida albicans* and *Candida tropicalis* growth, adhesion and biofilm development. *Mem Inst Oswaldo Cruz*, 111, 697-702.
- WEIDT, S., HAGGARTY, J., KEAN, R., COJOCARIU, C. I., SILCOCK, P. J., RAJENDRAN, R., RAMAGE, G. & BURGESS, K. E. 2016. A novel targeted/untargeted GC-Orbitrap metabolomics methodology applied to *Candida albicans* and *Staphylococcus aureus* biofilms. *Metabolomics*, 12, 189.
- WICKHAM, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*, Springer Publishing Company, Incorporated.
- WIMPENNY, J., MANZ, W. & SZEWZYK, U. 2000. Heterogeneity in biofilms. *FEMS Microbiology Reviews*, 24, 661-671.
- WINTER, M. B., SALCEDO, E. C., LOHSE, M. B., HARTOONI, N., GULATI, M., SANCHEZ, H., TAKAGI, J., HUBE, B., ANDES, D. R., JOHNSON, A. D., CRAIK, C. S. & NOBILE, C. J. 2016. Global Identification of Biofilm-Specific Proteolysis in *Candida albicans*. *MBio*, 7.
- WOLD, H. 1973. Nonlinear Iterative Partial Least Squares (NIPALS) Modelling: Some Current Developments. In: KRISHNAIAH, P. R. (ed.) *Multivariate Analysis—III*. Academic Press.
- WONG, E. H. J., NG, C. G., GOH, K. L., VADIVELU, J., HO, B. & LOKE, M. F. 2018. Metabolomic analysis of low and high biofilm-forming *Helicobacter pylori* strains. *Scientific Reports*, 8, 1409.
- WORLEY, B. & POWERS, R. 2013. Multivariate Analysis in Metabolomics. *Curr Metabolomics*, 1, 92-107.

References

- WU, S., LIU, Y., ZHANG, H. & LEI, L. 2019. The Pathogenicity and Transcriptome Analysis of Methicillin-Resistant *Staphylococcus aureus* in Response to Water Extract of *Galla chinensis*. *Evid Based Complement Alternat Med*, 2019, 3276156.
- WU, T., CEN, L., KAPLAN, C., ZHOU, X., LUX, R., SHI, W. & HE, X. 2015. Cellular Components Mediating Coadherence of *Candida albicans* and *Fusobacterium nucleatum*. *J Dent Res*, 94, 1432-8.
- XIANG, Y., YAN, B., YUE, B., MCNEFF, C. V., CARR, P. W. & LEE, M. L. 2003. Elevated-temperature ultrahigh-pressure liquid chromatography using very small polybutadiene-coated nonporous zirconia particles. *J Chromatogr A*, 983, 83-9.
- XU, X., HE, J., XUE, J., WANG, Y., LI, K., ZHANG, K., GUO, Q., LIU, X., ZHOU, Y., CHENG, L., LI, M., LI, Y., LI, Y., SHI, W. & ZHOU, X. 2015. Oral cavity contains distinct niches with dynamic microbial communities. *Environ Microbiol*, 17, 699-710.
- YAMAN, G., AKYAR, I. & CAN, S. 2012. Evaluation of the MALDI TOF-MS method for identification of *Candida* strains isolated from blood cultures. *Diagnostic Microbiology and Infectious Disease*, 73, 65-67.
- YITZHAKI, S., RESHEF, L., GOPHNA, U., ROSENBERG, M. & STERER, N. 2017. Microbiome associated with dentures malodour. *J Breath Res*.
- YU, G., WANG, L.-G., HAN, Y. & HE, Q.-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*, 16, 284-287.
- ZARNOWSKI, R., WESTLER, W. M., LACMBOUH, G. A., MARITA, J. M., BOTHE, J. R., BERNHARDT, J., LOUNES-HADJ SAHRAOUI, A., FONTAINE, J., SANCHEZ, H., HATFIELD, R. D., NTAMBI, J. M., NETT, J. E., MITCHELL, A. P. & ANDES, D. R. 2014. Novel entries in a fungal biofilm matrix encyclopedia. *MBio*, 5, e01333-14.
- ZAURA, E., BRANDT, B. W., PRODAN, A., TEIXEIRA DE MATTOS, M. J., IMANGALIYEV, S., KOOL, J., BUIJS, M. J., JAGERS, F. L., HENNEQUIN-HOENDERDOS, N. L., SLOT, D. E., NICU, E. A., LAGERWEIJ, M. D., JANUS, M. M., FERNANDEZ-GUTIERREZ, M. M., LEVIN, E., KROM, B. P., BRAND, H. S., VEERMAN, E. C., KLEEREBEZEM, M., LOOS, B. G., VAN DER WEIJDEN, G. A., CRIELAARD, W. & KEIJSER, B. J. 2017. On the ecosystemic network of saliva in healthy young adults. *ISME J*, 11, 1218-1231.
- ZHANG, W., LI, F. & NIE, L. 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology*, 156, 287-301.
- ZHANG, Z., YANG, J., FENG, Q., CHEN, B., LI, M., LIANG, C., LI, M., LI, Z., XU, Q., ZHANG, L. & CHEN, W. 2019. Compositional and Functional Analysis of the Microbiome in Tissue and Saliva of Oral Squamous Cell Carcinoma. *Frontiers in Microbiology*, 10, 1439.
- ZHAO, H. & EIDE, D. 1996. The yeast ZRT1 gene encodes the zinc transporter protein of a high-affinity uptake system induced by zinc limitation. *Proc Natl Acad Sci U S A*, 93, 2454-8.
- ZHAO, X., DANIELS, K. J., OH, S. H., GREEN, C. B., YEATER, K. M., SOLL, D. R. & HOYER, L. L. 2006. *Candida albicans* Als3p is required for wild-type biofilm formation on silicone elastomer surfaces. *Microbiology*, 152, 2287-99.
- ZHU, Z., WANG, H., SHANG, Q., JIANG, Y., CAO, Y. & CHAI, Y. 2013. Time course analysis of *Candida albicans* metabolites during biofilm development. *J Proteome Res*, 12, 2375-85.